

IFEDC-20194264

## 聚类方法在致密砂岩气层岩性识别中的应用

王英伟, 张景, 范希彬, 何金玉, 高阳

中国石油新疆油田 勘探开发研究院, 新疆 克拉玛依 834000

## The contrast of Several clustering methods in the tight sandstone gas reservoir lithology recognition

WANG Ying-wei,ZHANG Jing,FAN Xi-bin,HE Jin-yu,GAO Yang

Xinjiang Oilfield Company Exploration and Development Research Institution

Copyright 2019, Shaanxi Petroleum Society

This paper was prepared for presentation at the 2019 International Field Exploration and Development Conference in Xi'an, China, 16-18 October, 2019.

This paper was selected for presentation by the IFEDC Committee following review of information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the IFEDC Committee and are subject to correction by the author(s). The material does not necessarily reflect any position of the IFEDC Committee, its members. Papers presented at the Conference are subject to publication review by Professional Committee of Petroleum Engineering of Shaanxi Petroleum Society. Electronic reproduction, distribution, or storage of any part of this paper for commercial purposes without the written consent of Shaanxi Petroleum Society is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of IFEDC. Contact email: paper@ifedc.org.

### Abstract

Upper Paleozoic of L area is typical low porosity and low permeability of tight sandstone gas reservoirs. The lithology is fine and close, physical property is poorer. Lithologic control physical property, good or bad of physical property have key influence upon gas-bearing evaluation, reservoir classification evaluation. Lithology identification either for predicting gas yield area or to keep the accuracy of evaluation is very important. In this paper, based on principal component analysis. Application of K-mean clustering, and imulated annealing based on genetic for tight sandstone gas reservoir lithology by cluster analysis, imulated annealing based on genetic in tight sandstone gas reservoir lithology identification effect is better, lithology recognition correct rate as high as 74.42%, with good application effect.

**Keywords:** lithology identification; PCA; K-means clustering; imulated annealing based on genetic

### 摘要:

L 地区上古生界属于典型的低孔、低渗致密砂岩气藏。岩性致密、物性较差, 岩性控制物性, 物性的好坏对含气性的评价、储层的分类评价等有关键的影响, 所以岩性识别无论是对于预测天然气的高产区还是对保持评价精度都具有十分重要的意义。本文在主成分分析的基础上, 运用 K-均值聚类、基于遗传模拟退火算法聚类对致密砂岩气层岩性进行识别, 得到基于遗传模拟退火算法聚类在致密砂岩气层岩性识别中

的效果较好, 岩性识别正确率达到 74.42%, 应用效果良好。

**关键词:** 岩性识别;主成分分析;K-均值聚类;遗传模拟退火

岩性识别是地层评价、油藏描述的重要研究内容之一, 识别岩性最直接有效的方法是岩心分析, 但是考虑到油田上的生产效益, 深层钻井取心成本很高, 因此不可能在每口井中都取心, 因此利用间接的方法进行岩性识别成了必然。本文应用模式识别理论, 在综合分析测井与岩心资料的基础上, 研究 L 地区岩性变化特征, 结合具体情况, 在对测井曲线数据进行主成分分析后, 运用 K-均值聚类和基于遗传模拟退火算法聚类对 43 个岩性样本进行聚类, 通过对比得出基于遗传模拟退火算法聚类效果最好, 提高了致密砂岩气藏岩性识别的精度和效率。

## 一、主成分分析

主成分分析的原理为设有  $p$  个指标构成的  $p$  维随机向量  $X=(x_1, x_2, \dots, x_p)$ , 作指标的线性组合 (得综合指标) [1-3]:

$$Z_1=a_{11}x_1+a_{21}x_2+\dots+a_{p1}x_p \quad (1-1)$$

$$Z_i=a_{1i}x_1+a_{2i}x_2+\dots+a_{pi}x_p \quad i=1, 2, 3, \dots, p \quad (1-2)$$

$$Z_p=a_{1p}x_1+a_{2p}x_2+\dots+a_{pp}x_p \quad (1-3)$$

式(1-1),(1-2),(1-3)中,  $Z_i$  为主成分;  $a_{ij}$  为组合系数 (其中,  $j=1, 2, 3, \dots, p$ );  $x_i$  为第  $i$  个指标。组合系数  $a_{ij}$  根据下面的条件确定① $a_{1j}^2+a_{2j}^2+\dots+a_{pj}^2=1, j=1, 2, \dots, p$ ; ② $Z_1$  的方差在  $x_1, x_2, \dots, x_p$  的所有线性组合中最大;  $Z_1$  与  $Z_2$  不相关, 且  $Z_1$  的方差在  $x_1, x_2, \dots, x_p$  的所有线性组合中最大; 以此类推,  $Z_p$  与  $Z_1, Z_2, \dots, Z_{p-1}$  都互不相关, 且在  $x_1, x_2, \dots, x_p$  的所有线性组合中方差最大。

设  $X=(x_1, x_2, \dots, x_p)^T$  的协方差阵为  $D(X)$ ,  $D(X)$  的特征根为  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ,  $u_1, u_2, \dots, u_p$  为对应的标准特征向量, 第  $i$  个主成分为  $Z_i = u_i^T X, (i=1, 2, \dots, p)$ 。称  $\lambda_i / \sum_{k=1}^p \lambda_k$ ,  $(i=1, 2, \dots, p)$  为主成分  $Z_i$  的贡献率,  $\sum_{i=1}^m \lambda_i / \sum_{k=1}^p \lambda_k$  为  $Z_1, Z_2, \dots, Z_m$  的累积贡献率, 累积贡献率越大, 损失的数据信息就越少。根据累积贡献率判定主成分的个数, 一般要求累积贡献率不小于 80%。

## 二、聚类分析

### 2.1 K-均值聚类

K-均值聚类(K-means clustering)是 J.B. Mac Queen 很早提出的一种无监督的实时聚类算法, 在最小化误差函数的基础上将数据划分为预定的类数  $C$ 。K-均值聚类算法的一般步骤[4]:

#### ①初始聚类中心

根据具体问题, 凭经验从样本集中选出  $C$  个比较合适的样本作为初始聚类中心。b. 将全部样本随机地分成  $C$  类, 计算每类样本均值, 将样本均值作为初聚类中心。

#### ②初始聚类

按就近原则将样本归入各聚类中心所代表的类中。b. 取以样本, 将其归入与其最近的聚类中心的那一类中, 重新计算样本均值, 更新聚类中心。然后取下一样本, 重复操作, 直至所有样本归入相应类中。

#### ③判断聚类是否合理

采用误差平方和准则函数判断聚类是否合理, 不合理则修改分类。循环进行判断、修改直至达到算法终止条件。

### 2.2 基于遗传模拟退火算法

将模拟退火算法与遗传算法相结合 (SAGA) 用于聚类分析, 由于模拟退火算法和遗传算法可以互相取长补短, 因此有效地克服了传统遗传算法的早熟现象, 同时根据聚类问题的具体情况设计遗传编码方式及适应度函数, 使该算法更有效、快速地收敛到全局最优解。

基于遗传模拟退火算法的聚类步骤[5]如下:

- ①初始化控制参数:种群个体大小  $sizepop$ ,最大进化次数  $MAXGEN$ ,交叉频率  $P_c$ ,变异概率  $P_m$ ,退火初始温度  $T_0$ ,温度冷却系数  $k$ ,终止温度  $T_{end}$ 。
- ② 随机初始化  $c$  个聚类中心,并生成初始种群  $Chrom$ ,对每个聚类中心计算各样本的隶属度,以及每个个体的适应度值  $f_i$ ,其中  $i=1,2,\dots,sizepop$ 。
- ③ 设循环计数变量  $gen=0$ 。
- ④对群体  $Chrom$  实施选择、交叉、变异等遗传操作,对新产生的个体计算  $c$  个聚类中心、各样本的隶属度,以及每一个体的适应度值  $f_i'$ 。若  $f_i' > f_i$ ,则以新个体替换旧个体; 否则,以概率  $P=\exp((f_i-f_i')T)$
- ⑤若  $gen<MAXGEN$ ,则  $gen=gen+1$ ,转至步骤③; 否则,转至步骤⑥。
- ⑥若  $T_i<T_{end}$ ,则算法成功结束,返回全局最优解; 否则,执行降温操作  $T_{i+1}=kT_i$ ,转至步骤③。

三、 3 实例应用

L 地区储层的岩石类型以石英砂岩、岩屑砂岩、岩屑石英砂岩三种岩石类型为主。基于 8 口井 43 块岩心薄片资料分析与现场岩心观察,其中石英砂岩、岩屑砂岩、岩屑石英砂岩各分别为 11、15、17 块。

3.1 主成分分析

测井曲线 GR, DEN, CNL, AC, RD, RS 是反应岩性特征变量,对这 6 个变量进行主成分分析。表 3-1 是主成分分析提取的结果,从表中可以看出当有 3 个主成分时,累计贡献率已经达到 86.442%,超过了 80%,所以只需要 3 个主成分就可以满足要求。

表 3-1 主成分分析提取结果

成分	特征值	方差	累计
1	2.199	36.646	36.646
2	1.815	30.257	66.903
3	1.172	19.538	86.442
4	0.460		
5	0.285		
6	0.068		

表 3-2 成份矩阵

成分	$Z_1$	$Z_2$	$Z_3$
RS/ $\Omega\cdot m$	0.937	-0.109	0.265
RD/ $\Omega\cdot m$	0.925	-0.098	0.223
DEN/ $g\cdot (cm^3)^{-1}$	-0.114	0.890	-0.46
AC/ $\mu s\cdot m^{-1}$	-0.504	-0.684	0.002
CNL/%	-0.340	-0.383	0.799
GR/API	-0.290	0.622	0.642

提取了 3 个主成分  $Z_1$ 、 $Z_2$ 、 $Z_3$ , 它们的成份矩阵如表 3-2 所示,根据成份矩阵可以得到  $Z_1$ 、 $Z_2$ 、 $Z_3$  这 3 个主成分的表达式:

$$Z_1=0.937\times RS+0.925\times RD-0.114\times DEN-0.504\times AC-0.340\times CNL-0.290\times GR \quad (3-1)$$

$$Z_2=-0.109\times RS-0.098\times RD+0.890\times DEN-0.684\times AC-0.383\times CNL+0.622\times GR \quad (3-2)$$

$$Z_3=0.265\times RS+0.223\times RD-0.460\times DEN+0.002\times AC+0.799\times CNL+0.642\times GR \quad (3-3)$$

式(3-1),(3-2),(3-3)中,GR 为自然伽马,API; DEN 为密度, $g/cm^3$ ; CNL 为中子测井,%; AC 为声波时差, $\mu s/m$ 。RD 为深侧向电阻率, $\Omega\cdot m$ ; RS 为浅侧向电阻率, $\Omega\cdot m$ 。

下文的几种聚类结果和岩性实际分类对比图中,类别标签分别取 1,2,3。

1=石英砂岩, 2=岩屑砂岩, 3=岩屑石英砂岩。

3.2 K-均值聚类

表 3-3 K-均值聚类的初始聚类中心

变量	初始聚类中心		
	1	2	3
$Z_1$	-119.19	1082.17	622.75
$Z_2$	-153.35	-252.02	-199.46
$Z_3$	48.60	331.66	237.12

表 3-4 K-均值聚类的最终聚类中心

变量	最终聚类中心		
	1	2	3
$Z_1$	-3.12	1082.17	562.19
$Z_2$	-131.33	-252.02	-187.17
$Z_3$	76.17	331.66	223.69

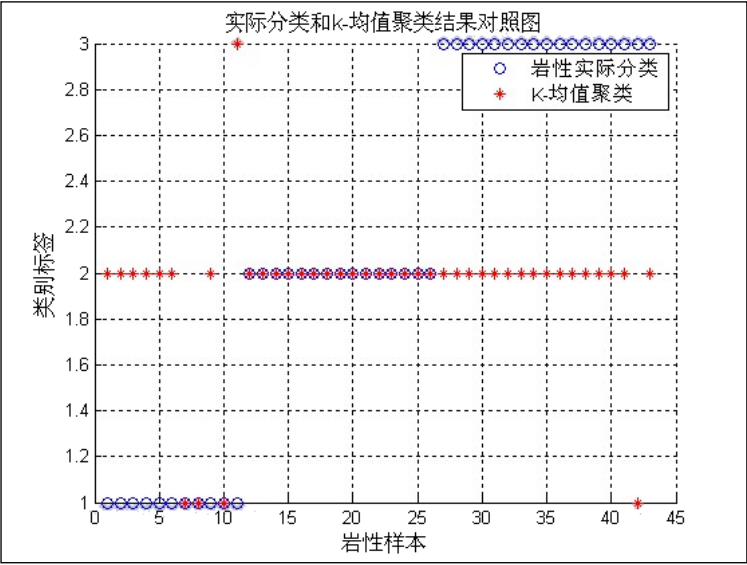


图 3-1 K-均值聚类岩性样本结果和实际分类结果对照

从图 3-1 可以看出 K-均值聚类预测结果和岩性的实际分类结果有很大差距：43 个岩性数据中只有 18 个聚类正确，正确率为 41.86%。

### 3.3 基于遗传模拟退火算法聚类

冷却系数  $q=0.8$ ,初始温度  $T_0=100$ ,终止温度  $T_{end}=1$ ,最大遗传代数  $MAXGEN=100$ 。基于遗传模拟退火算法的最终聚类中心如表 6 所示。

根据主成份分析得到的 3 个主成分对三种岩性的数据运用基于遗传模拟退火算法的聚类方法进行聚类,结果如图 3-3 所示。

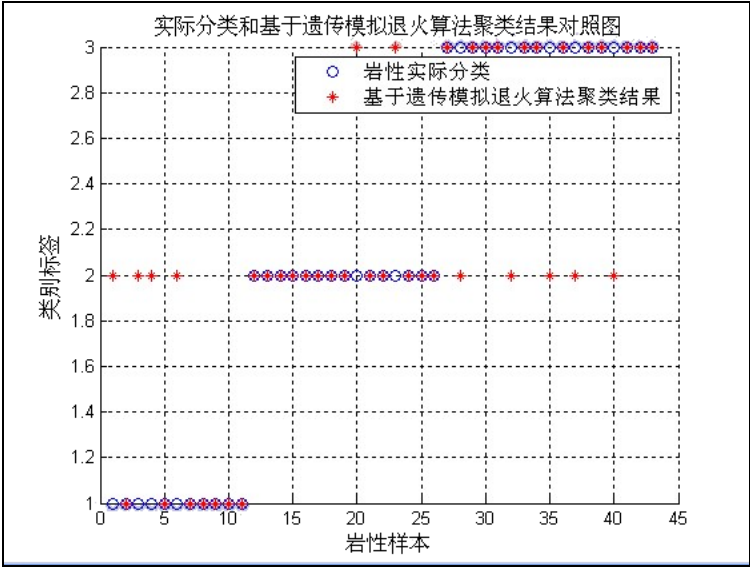


图 3-2 基于遗传模拟退火聚类岩性样本和实际分类结果对照

从图 3-2 可以看出基于遗传模拟退火算法的聚类效果较好：43 个岩性数据中有 32 个聚类正确，正确率为 74.42%。

综上所述，K-均值聚类和基于遗传模拟退火聚类两种聚类方法的正确率分别是 41.86%和 74.42%。基于模拟退火聚类算法跟另两种算法相比具有一定的优势,不容易陷入局部极值，在岩性识别中效果较好。

## 四、 结语

通过模式识别中的主成分分析对反应岩性特征的测井曲线进行处理，得到三个反应岩性特征的综合参数。在此基础上，运用 K-均值聚类和基于遗传模拟退火算法聚类对岩性进行聚类分析，通过与实际岩性类

别进行对比,得到了基于遗传模拟退火算法的聚类在岩性分类中效果最好,而且证实了该方法是有效可靠的。

## 参考文献

- [1] 谢明霞,郭建忠,陈科等.综合国力主成分分析与模糊综合评价模型[J].地理与地理信息科学,2010,26(6):106-108.
- [2] 张崇甫,陈述云.成分数据主成分分析及其应用[J].数理统计与管.1996,15(3):11-14.
- [3] 杨中荣,毛广运,臧桐华等.用 SAS 和 SPSS 软件进行主成分分析[J].中国卫生统计,2009,(2):212-213.
- [4] 杨淑莹.模式识别技术与智能计算[M].北京:电子工业出版社,2008:253-256.
- [5] 史峰,王辉,郁磊等.Matlab 智能算法 30 个案例分析[M].北京:北京航空航天大学出版社,2011:188-196.

[作者简介]王英伟,(1985-),男,助理工程师,2012年毕业于长江大学获得硕士学位,研究方向为测井解释、测井地层评价、油藏描述,现在中石油新疆油田公司勘探开发研究院工作。