

文章编号: 0253- 9985(2004) 01- 0102- 04

模式识别在地层划分中的应用

陈清华^{1,2}, 章凤奇², 孙述鹏³, 刘池阳¹

(1. 西北大学地质系, 陕西西安 710069; 2. 石油大学石油资源科学系, 山东东营 257061;

3. 石油大学图书馆, 山东东营 257061)

摘要: 模式识别就是研究已知的模式, 并对模式进行分类, 利用计算机技术, 将待识别模式进行归类, 达到分类的目的。模式识别的过程可概括为: 首先, 应用傅立叶频谱分析、回归分析及多点平滑法进行数据预处理; 其次, 采用旋回地层学理论, 计算已知钻井地质数据的频谱图, 利用米兰柯维奇旋回组分进行学习, 提取特征, 建立模式分类; 最后, 用已建立的模式类对其他钻井进行地层划分。运用模式识别方法, 对大港油田 40 多口钻井进行了地层划分, 划分结果与原地质分层结果具有很好的相关性, 而且精度更高。

关键词: 模式识别; 地层划分; 旋回地层学

中图分类号: TE122. 1 文献标识码: A

Application of pattern recognition in stratigraphic division

Chen Qinghua^{1,2} Zhang Fengqi² Sun Shupeng³ Liu Chiyang¹

(1. Department of Geology, Northwest University, Xi'an, Shaanxi; 2. Petroleum Resource Science Department, University of Petroleum, Dongying, Shandong; 3. Library of University of Petroleum, Dongying, Shandong)

Abstract: Pattern recognition is to subsume undetermined patterns with computer technologies based on study and classification of the known patterns. Pattern recognition can be performed in three steps. First, preprocess data through Fourier analysis, regression analysis and multi-point smoothing; then calculate frequency spectrum of the known wells' geological data by using cyclical stratigraphy and perform feature study and extraction and pattern classification with Milankovitch cyclic components; and finally classify strata in other wells by using the established patterns. Stratigraphic division has been performed in 40 wells in Dagang oilfield with the pattern recognition method and the results are in good agreement with the original ones and have better accuracy.

Key words: pattern recognition; stratigraphic division; cyclical stratigraphy

地层划分是地质学研究最基本问题, 它直接影响着沉积学、构造地质学及其他相关应用地质学的研究。因此, 地层学研究一直是人们关心的热点问题。目前, 地层学研究方法与技术很多, 例如同位素法^[1]、古生物法^[2]、古地磁法^[3]、旋回地层学分析^[4]等, 但通过计算机模式识别技术并结合地质分析形成系统分析软件的并不多。本文通过与大港油田的合作攻关项目“大港油田旋回地层学分析软件开发”的研究, 系统介绍模式识别方法在地层划分中的应用。

1 技术简介

1.1 概念^[5]

对于模式识别的理解不同人有不同的看法, 但总的说来不外模式、模式类和模式识别几个概念。其中模式可理解为个别事物或现象, 模式类就是具有某些共同特性的模式的集合。模式识别就是研究一种自动技术, 依靠这种技术, 机器将自动地把待识别模式分配到各自的模式类中去。

目前, 模式识别主要应用于文字和符号识别、图像识别、声音识别、生物医学等方面, 在地质领

域应用较少。

1.2 一般过程

模式识别的全过程一般是首先对被识别对象(样本)进行数字化和预处理,滤除样品中的干扰、噪声,并人为地突出有用信号,然后作特征抽取或基元抽取(视分类方法而定),并对其分类。若识别效果差则可返回到预处理重新进行特征抽取(或改进处理方法)。

2 应用

本软件分两步来实现,第一步是机器学习和模式特征提取阶段。这一阶段的主要任务是通过已知地层分层和年代的井的学习获得模式特征,为下一步作准备。第二步是对未知井进行地层划分,它是在第一步的基础上完成的。系统整个过程是以旋回地层学为主要理论依据。以下详细说明如何应用模式识别技术与地质分析相结合的方法,解决地层划分的问题。

2.1 采样数据(伽马测井数据)预处理

被识别对象是自然伽马测井数据,因为它本身已经是数字化信息,所以不再需要进行数字化这一过程,而直接转到预处理过程。对于采样数据进行预处理,我们应用了多种模式识别技术,主要有下面几种。

2.1.1 傅里叶频谱分析^[6]

用傅里叶变换进行频谱分析,来压制超高频和极低频的信号,突出深层反射的地震响应记录,这种方法在石油勘探领域应用十分广泛。经过傅里叶变换,把在时间域里看起来很复杂的多频率叠加信号分解成不同频率的正弦分量,将空间域变换到了频率域。由于计算机技术的需要和实际数据分析长度的限制,我们通常采用离散傅里叶变换,即对有限长度的离散时间序列 $x(n)$ 进行傅里叶变换,其定义为

$$X[m] = \sum_{n=0}^{N-1} x[n] W_N^{mn}$$

其逆变换表达式为

$$X[n] = \frac{1}{N} \sum_{m=0}^{N-1} x[m] W_N^{mn}$$

其中 $m = 0, 1, \dots, N-1$, $W_N = e^{j(2\pi/N)}$

离散傅里叶变换的频谱是连续的周期谱。由于计算机只能记录离散的点,因此,我们用计算机进行离散傅里叶变换(DFT)和快速离散傅里叶变

换(FFT),得到的频谱都只是对连续变换频谱的离散采样。为了更精确的表示实际信号的特点,就需要考虑采样间隔、序列长度等问题。我们采用了基-2快速傅里叶变换算法,即处理数据序列的长度为 2^n , n 为整数。对于实际分析数据个数不为 2^n 的地层数据段,我们采用两边补零的方法。

2.1.2 回归分析

地层原始沉积过程中存在很多因素,除了高频旋回地层学研究的短周期控制因素以外,还有长期变化趋势,如构造升降的影响,以百万年计的长期天文变化的影响,因此我们得到的数据可能是具有长期变化趋势的。在做高频旋回分析时应该首先把长周期变化的影响去掉。在所研究地层的区间内,长期变化是接近线性的。所以,我们采用线性回归分析来得出回归方程,把总体趋势求出来。

最小二乘法进行一元线性回归是比较简单可行的方法,它对于接近线性变化的数据比较适用。

2.1.3 数据序列的光滑处理——低通滤波^[7]

测井序列中夹杂很多相对目的信号要高很多的频率成分,因此要进行低通滤波,把高频成分滤掉。低通滤波可以在频率域进行,也可以在时间域进行。频率域的低通滤波是把要滤去的高频部分频率的实部和虚部分别清零,也就相当于在频率域里乘上滤波函数的谱。所有的滤波过程都可以在时间域里进行,其方法是把要分析的信号 $f(t)$ 和滤波函数 $h(t)$ 进行褶积运算(卷积运算)

$$\begin{aligned} f(t) * h(t) &= \int_{-\infty}^{\infty} f_1(\tau) f_2(t - \tau) d\tau \\ &= \int_{-\infty}^{\infty} f_1(t - \tau) f_2(\tau) d\tau \end{aligned}$$

时间域的卷积和频率域的乘积是一对等价变换关系

$$f(t) * h(t) \leftrightarrow F(\omega) H(\omega)$$

在时间域里进行的多点平均光滑方法也相当于低通滤波,多点移动平均法,也叫多点平滑法,都是采用奇数个点,平滑点数采用的越多,滤掉的高频成分越多,平均效应越强,对时间域里多采用9点,11点,13点平均平滑方法。根据平滑算法的频率响应曲线(图1),可以看出平滑次数越多,其频率响应的通带越窄,过滤掉的高频成分越多,频谱分辨率越高。为了减小运算量又保证处理效果,采用11点或13点6次平滑比较合理。

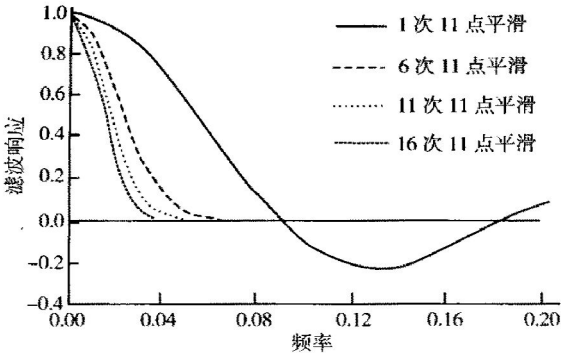


图 1 11 点平滑的频率响应

Fig. 1 Frequency response of 11 points smoothing

为了更加准确的获取米兰柯维奇旋回,除了进行光滑、去趋势,还要对频谱进行光滑处理以消除高频现象造成的影响。

2.2 特征提取和模式分类

2.2.1 模式类的确定

根据旋回地层学理论,斜率、岁差旋回都分别有两个分量,按周期由长到短分别记为 C, D, E, F 。各周期对沉积地层的旋回控制作用强弱不同,有些地层发育 C, F 相应的旋回,有的地层发育 C, E, F 相应的旋回等^[4, 7~12]。我们利用偏心率(记为 A)与这 4 个周期的任意组合(共 15 种)的比值分别进行旋回识别^[13, 14]。旋回地层学方法主要研究第三纪、白垩纪、侏罗纪、三叠纪、二叠纪、石炭纪、泥盆纪、奥陶纪、寒武纪 10 个大的地质年代。10 种地质年代和 15 种周期的组合又形成 150 个模式类(当然也可将地层划分得更细从而得到更多的模式类),模式类的主要特征就是地质年代和周期的组合形式。

2.2.2 模式分类

当模式类确定之后,我们就可通过模式识别分类来确定属于某一区块的模式类。具体的方法就是通过对某一区块已知的若干井进行学习处理,从这些已知井中抽取出这一区块的共同特征,然后区分出不同的模式类(图 2)。

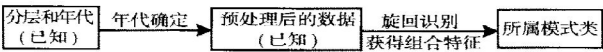


图 2 模式类确定过程

Fig. 2 Procedure of pattern determination

根据旋回地层学理论,旋回识别包括对预处理后的数据进行频谱分析、旋回提取和确定地质

年代。确定地质年代时分别采用 15 种周期组合,求出各个谱峰对应周期的比值,如果有一组比值接近米兰柯维奇旋回的周期比,则可以认为它们是天文轨道参数作用产生的旋回。根据这种结果可以进行旋回识别,经过反复地计算和识别获得决定该区块的米兰柯维奇旋回天文周期组分。下面给出通过这一阶段学习和归类得到的某一区块的若干模式类(表 1),表 1 中 ω 代表模式类, ω 的个数一般为 10 个。

表 1 某一区块地层的若干模式类

Table 1 Several patterns of strata in a certain block

模式类	地质年代	斜率、岁差的可能组合
ω_1	第三系	C, F
ω_2	白垩系	E, F
ω_3	侏罗系	B, C, E
ω_4	三叠系	B, E
ω_5	二叠系	C, E

2.3 地层划分

当某一区块的模式类决定之后,即可利用它进行地层划分。旋回地层学理论认为米兰柯维奇旋回在地层中保存状况受地球轨道参数的间接控制,地层沉积的水体环境随轨道参数的变化而变化,沉积物的各种特性变化具有与米兰柯维奇天文旋回相同的周期。对于连续沉积的地层,其频率成分是渐变的,米兰柯维奇旋回的各个天文周期的组分均可进行识别。而对于间断面附近的频谱来说,由于旋回信息没有被完整的保留下来,造成米兰柯维奇旋回的个别参数,甚至 3 个参数的 5 种周期均无法识别。表现在频谱图上也即没有对应岁差、斜率、偏心率中某一个或几个参数周期的峰值^[15]。因此,利用旋回保存程度可以判断沉积间断面。具体方法是选定合适大小的分析窗口,计算每个窗口内的地质数据(已经过预处理)的频谱图,利用比值法识别米兰柯维奇旋回组分;然后窗口移动一定步长(步长一定要小于二分之一窗长),分析下一段数据,这就是滑移窗技术,据此得出频率变化的频带图。利用识别的米兰柯维奇旋回,做出具有该区块模式类特征的旋回发育曲线,对于旋回间断面存在的深度位置赋以特定值(如特大值),旋回发育的深度赋以另一种值(如特小值 0),这样就得出若干条(曲线条数与该区块模式类个数相同)旋回发育状况曲线,据各条曲线旋回

间断点的数目判断小层间断面的存在与否, 从而进行地层划分。

2.4 应用实例与效果分析

采用自编软件, 对大港油田徐杨桥- 黑龙村潜山构造带、歧口凹陷、盐山凹陷等构造单元的徐13井、徐14井、庄古1井、庄63井、扣11、扣21、新港508-4等40余口井进行了地层划分, 以便于进行分层结果的可比性检验。图3示应用本软件对盐14井的分层结果。研究区为砂泥岩剖面, 自动分层研究采用了对泥质反应敏感的伽马测井数据, 图3中频带图栏内横线为分层界限, 由软件自动生成, 其左侧细横线为原地质分层。

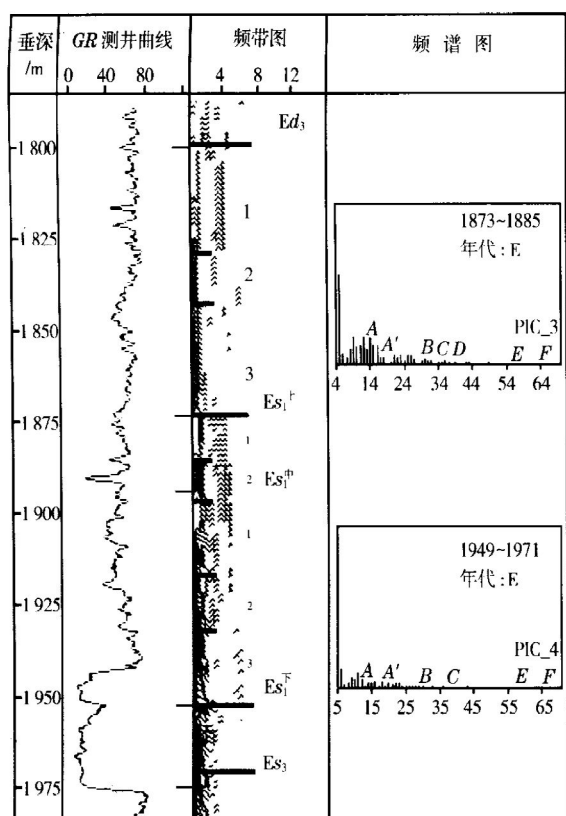


图3 大港油田盐14井米兰柯维奇旋回法分层图

Fig. 3 Stratigraphic division with Milankovitch

Cyclical method in Yan 14 well in Dagang oilfield

盐14井沙河街组一段原地质分层分为上、中、下3个砂层组, 利用该方法对盐14井进行自动分层的结果识别出了这3个砂层组的界限, 并进一步将上砂层组细分为3个次级地层单元, 中砂层组细分为两个次级地层单元, 下砂层组细分为3个次级地层单元, 分层结果比原地质分层细。实际上, 该方法所划分的地层单元为小于百万年的

高频旋回, 相当于层序地层学中的4~5级沉积旋回。

40余口井的对比结果表明, 模式识别技术划分的结果和原地质分层结果具有很好的相关性, 并在分层精度上更加精确。反映该方法地质吻合率高和实用性强的特点。

依据数据中蕴含的频率信息, 仅需要测井数据即可进行模式识别并进行分层。可以适用于砂泥岩地层、碳酸盐岩地层, 对于缺乏地质标志的中生代地层更显示其优势。研究证明, 除了测井数据以外, 其他包含旋回信息的数据, 如地球化学数据、岩性数据、地层厚度数据等均可应用软件进行分析, 且分层速度快。反映软件具有适应性强、效率高的特点。

测井数据是地层沉积物性变化的综合反映, 每种测井数据对地层沉积环境反映程度不同, 但它们都具有数据量大、采样密集、纵向分辨率高的优点, 这种数据可以直接用来处理, 无需再采样和读取数值。对于不同岩性的地层, 应该采用不同的测井数据进行模式识别和分层, 如砂泥岩剖面, 选用自然伽马测井数据, 碳酸盐岩剖面, 选用倾角测井等, 不论采用哪种测井数据, 模式识别和分层的具体方法、步骤基本一致。经现场实验检验, 对于中国东部第三纪地层, 选取自然伽马测井数据处理精度最高、效果最好。

3 结语

模式识别技术与石油地质研究相结合的研究方法, 可以很好的解决石油领域的一些问题。随着计算机理论的进一步发展, 我们相信模式识别将会更广泛地应用于石油领域, 各种智能化程度很高的应用分析软件将会出现, 从而更好的服务于石油工业。

参 考 文 献

- 1 叶得泉, 钟筱春, 唐文松, 等. 稳定同位素地层学和事件地层学在白垩系-第三系界线划分中的应用[J]. 微体古生物学报, 1994, 11(2): 141~152
- 2 陈中强, 钟石兰, 杨少坤, 等. 微古生物在高分辨层序地层学中的应用——以珠江口盆地惠州凹陷层序地层分析为例[J]. 微体古生物学报, 1998, 15(2): 134~143
- 3 杨振宇, 马醒华, 孙知明, 等. 华北地块显生宙古地磁视极移曲线与地壳运动[J]. 中国科学(D辑), 1998, 28(1): 44~56

(下转第114页)

计算: a) 多元线性回归分析; b) 在没有进行偏相关分析优选参数的前提下, 使用了所有的地层参数和施工参数进行 GM(1, 10) 分析。在进行方法对比时, a) 以试井解释的裂缝导流能力和裂缝半长值为标准进行误差计算; b) 以压后和压前的无阻流量比值为增产倍数, 进行误差计算。

我们根据所建立 GM(1, 7) 预测模型, 对另外 50 余口压裂井进行了预测, 并且和压力恢复试井、产能测试结果对比, 进行误差计算, 各方法的计算误差对比如表 1。

Table 1 Comparison of outcomes of various methods			
预测方法	导流能力 误差均值/ %	裂缝半长 误差均值/ %	增产倍数 误差均值/ %
偏相关+ GM(1, 7)	2.5	3.6	6.3
多元线性回归	15.2	25	35.8
GM(1, 10)	6.9	5.9	8.5

4 结论

(1) 在进行偏相关分析后, 起到了优选参数的作用: 本文采用偏相关+ GM(1, 7) 模型优选出和导流能力相关的参数: 层厚、声波时差、泥质含量、加砂量、施工排量、前置液量, 共计 6 个影响变量, 排除了 3 个不相关因素的干扰, 其拟合的精度和直

接应用灰色模型 GM(1, 10) 相比, 有较大提高。

(2) 应用多元线性回归进行的导流能力拟合误差为 15.2%, 比其他两种方法的误差都要大得多, 证明导流能力等压裂效果参数和其影响参数为非线性关系, 多元线性回归方法不适合于新场气田压裂效果的拟合和预测。

(3) 在进行增产倍数预测时, 几种方法的预测误差均值都较大, 主要是因为增产倍数还和表皮系数等无法预先估算的因素有关, 其影响因素远比裂缝导流能力和裂缝半长的影响因素复杂。

参 考 文 献

1 冯文光. 污染引起的渗流异常机理与 SLUG 试井分析原理[M]. 成都: 成都科技大学出版社, 1999

2 陈伟, 段永刚. 水平裂缝压裂井试井分析[J]. 油气井测试, 2000, 9(3): 47~ 50

3 卢纹岱. SPSS for Windows 统计分析[M]. 北京: 电子工业出版社, 2000

4 徐忠祥, 吴用平. 灰色系统理论与矿床灰色预测[M]. 武汉: 中国地质大学出版社, 1993

5 胥泽银, 郭科. 多元统计方法及其程序设计[M]. 成都: 四川科学技术出版社, 1999

6 吴志均, 唐红君. 川西致密砂岩气藏压裂效果的试井解释评价[J]. 油气地质与采收率, 2002, 9(6): 52~ 54

7 张莉, 岳乐平. 鄯善油田地应力、裂缝系统与油田开发[J]. 石油与天然气地质, 1999, 20(4): 330~ 332

8 吉懋利. 水力压裂技术新进展[M]. 北京: 石油工业出版社, 1995

(上接第 105 页)

4 吴智勇, 姜衍文. 全球旋回地层学述评[J]. 石油与天然气地质, 1996, 17(1): 15~ 21

5 沈清, 汤霖. 模式识别导论[M]. 北京: 国防科技大学出版社, 1995

6 陈行禄, 秦永元. 信号分析与处理[M]. 北京: 北京航空航天大学出版社, 1993

7 石万忠, 李举子, 李梦溪, 等. 时- 频分析在高分辨率层序地层学中的应用[J]. 石油与天然气地质, 2001, 22(3): 235~ 237

8 Hays J D, Imbrie J, Shackleton N J. Variation in the Earth's Orbit: pacemaker of the ice ages[J]. Science, 1976, 194(4 270): 1 121~ 1 132

9 Erger A L. Spectrum of climate variations and their causal mechanisms [J]. Geophisics Surveys, 1979, 3: 351~ 402

10 Olsen H. Astronomical forcing of meandering river behaviour: Milankovitch cycles in Devonian of East Greenland[J]. Palaeogeography, Palaeoclimatology, Palaeoecology, 1990, 79: 99~ 115

11 Kashiwaya K, Ryugo M. Long term climato-limnological oscillation during the past 2.5 million years printed in Late Baikal sediments[J]. Geophysical Research Letters, 1998, 25(5): 659~ 662

12 姜衍文, 朱忠德, 肖传桃, 等. 全息地层学——新的挑战[J]. 石油与天然气地质, 1995, 16(2): 110~ 118

13 Yang C S, Baumfalk Y A. Milankovitch cyclicity in the Upper Rotliegend Group of the Netherlands Offshore[J]. Spec Publs Int Ass Sediment, 1994, 19: 47~ 61

14 齐永安, 王润怀, 潘结南. 米兰柯维奇旋回与高频波动周期的识别[J]. 焦作工学院学报(自然科学版), 2000, 19(2): 81~ 85

15 郑晓武, 邓宏文, 徐怀大, 等. 测井高分辨率层序地层自动划分技术与应用尝试[J]. 石油与天然气地质, 1999, 20(4): 357~ 360