

分级特征提取在中子寿命深度校正中的应用^①

吴雅娟¹, 徐亚军¹, 赵鹏²

¹(东北石油大学 计算机与信息技术学院, 大庆 163318)

²(大庆油田测试技术服务分公司, 大庆 163453)

摘要: 随着油田进入开发中后期, 中子寿命测井得到了广泛的应用. 针对中子寿命测井曲线存在的深度误差, 对两条测量不同地层信息的曲线进行对比校正, 利用测井曲线顺序性的特点, 对测井曲线进行分级特征提取并进行校正, 提出了基于分级特征提取的测井曲线校正的方法. 借鉴人工视觉校正的经验, 利用曲线段的最大值位置、尖峰数目这两个特征先对曲线段进行分类匹配, 然后利用最大值上下厚度比、相对斜率这两个特征进行相似性判断最终实现分级特征提取. 实验结果表明该方法能够快速有效的对测井曲线进行校正.

关键词: 中子寿命; 深度校正; 顺序性; 分级特征提取

Hierarchical Feature Extraction in the Application of Neutron Lifetime Depth Correction

WU Ya-Juan¹, XU Ya-Jun¹, ZHAO Peng²

¹(School of Computer & Information Technology, Northeast Petroleum University, Daqing 163318, China)

²(Logging & Testing Services Company, Daqing Oilfield CO.LTD, Daqing 163453, China)

Abstract: Neutron lifetime logging has been widely used, as the oilfield enters the mid and late development. Aiming at the existence of neutron lifetime logging curve depth error, contrast correction curve of two different stratigraphic information, and take advantage of the characteristics of logging curve sequential, for logging curve of hierarchical feature extraction and calibration. Logging curve correction method is presented based on the hierarchical feature extraction. Using the maximum position, spike number to match and categorize, and then using the maximum thickness ratio, and relative slope characteristics of the curve segment to judge similarity, and Finally realize the hierarchical feature extraction, these all based on the experience of the artificial vision correction. The experimental results show that the method can quickly and efficiently correct the logging curves.

Key words: neutron life time; depth correction; sequential; hierarchical feature extraction

1 引言

油田进入开发中后期, 已进入高含水期, 油水状况跟开发前已经有了很大变化, 而且由于油田构造复杂、油藏埋藏深、矿化度变化范围大, 区块特征异常显著, 层间非均质性差, 高含水层及相对较差油层交互, 常规剩余油监测技术不能准确识别出水层及油层水淹状况. 而中子寿命测井不受岩性、温度影响, 而且分辨率比较高, 能够分辨 0.3 m 的薄层, 能够准确地分出油层. 所以中子寿命测井在油田中后期得到了广泛应用^[1].

中子寿命测井对于测井数据的精度要求很高, 而且要正确识别出薄油层更对测井曲线的深度精度有很高的要求. 但是由于测量产生的深度误差是不能避免的而且往往误差较大, 要想根据中子寿命测井准确分出层位, 必须对测井曲线进行深度校正, 本文将对中子寿命测井的自动化深度校正进行研究.

石油领域的测井数据是以等距方式采集一组物理量的数据^[2], 并以关系模式形式记录, 常常以一组曲线形式直观显示, 称为测井曲线.

目前在测井曲线的深度校正主要运用的方法是曲

① 基金项目:国家自然科学基金(61170132);国家重大专项(2011ZX05020-007);黑龙江省教育厅科学技术研究项目(12521055)

收稿时间:2013-08-05;收到修改稿时间:2013-09-02

线的相似性, 曲线的相似性对比研究中, 有基于距离表示的算法, 有基于形态表示的算法^[3]. 而对曲线相似性的判断, 有比如文献[4]的提出的形态特征(时间序列的斜率表示及斜率距离)的曲线相似性搜索, 文献[5]提出的动态规划字符串匹配算法在曲线对比中的应用. 文献[6]提出的最大梯度相似匹配算法. 目前判断曲线相似性的方法大都是比较的曲线段的斜率、厚度或者峰值等. 而由于中子寿命测井校正的特殊性, 比较的是两条测量不同地层信息的曲线. 而且待校正的曲线存在拉伸或者缩小的情况, 甚至会有异常点的情况. 所以上面所提到单纯的曲线相似性的特征提取和判断方法对于中子寿命测井的深度校正应用有一定的局限性. 本文综合人工校正的经验, 根据测井曲线顺序性(测井曲线是不会出现相互交织的情况的, 深度大的曲线只可能在深度小的曲线下面, 不会相互交织)的特点^[7], 提出了一种基于中子寿命测井的分级特征提取的深度校正方法. 该方法将连续的数值曲线转换为数字序列, 简化了对两条测量不同地层信息曲线校正的特征提取的过程, 能够很大地提高校正的效率.

2 基于重要点分段

数值形式表示的曲线不便于对比的描述. 为此, 本文将数值曲线序列转换成离散的、相对抽象的数字序列, 每一种数字代表一段曲线也代表一种基本的、相对独立的变化趋势.

对曲线分段的方法有多种, 有基于概率统计的分段方法, 有对截面曲线进行分类的^[8,9]. 基于重要点的分段方法认为曲线变化是由一系列的事件引起的, 如在地质过程中, 水进水退会引起曲线幅度的变化, 按事件发生的重要程度对曲线进行划分^[10].

重要极小/(大)点定义: 给定常量 R 和测井序列 $\{X_1, \dots, X_n\}$, 对于 $1 < m < n$, 存在下标 i 和 j , 且 $1 \leq i < m < j \leq n$, 有 X_m 是 X_1, \dots, X_j 中的最小/(大)值; $X_i/X_m \geq R$, 且 $X_j/X_m \geq R$ 成立. 其中 R 是可控选取的参数, R 值大则被选中的相对重要点少, 曲线分段就越粗; 反之, 分段就越细. 因此, 通过选择 R , 可以在不同的精细程度上进行曲线分段. 基于重要点的分段方法在一定程度消除了仪器、背景等噪声的影响, 保持了序列变化的主要特征模式. 如图 1, 本文对测井曲线的分段图, 而本文的最终目的就是找到两条曲线对应的曲线段, 然后再进行深度校正.

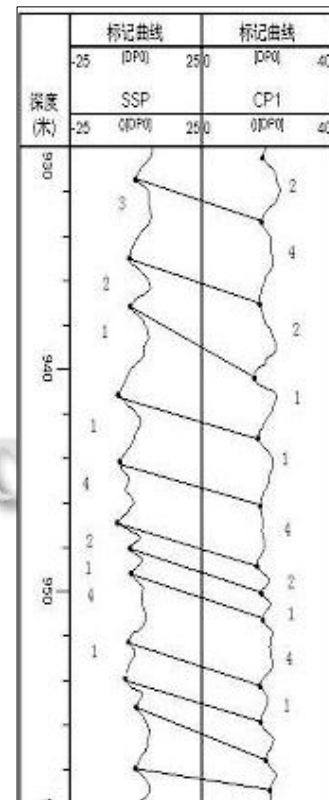


图 1 测井曲线分段图

3 分级特征提取

中子寿命的曲线校正基于 SSP 曲线对 $CP1$ 和 $CP2$ 进行校正. (SSP 为电磁流量曲线, $CP1$ 、 $CP2$ 为中子寿命测井的替前、替后曲线).

利用测井曲线的顺序性, 有时并不需要对整条曲线进行详细的相似性判断, 只要查找出两条曲线的曲线段相互匹配的位置就能进行校正. 如图一所示, 我们只要找到图中的对应关系, 就能很方便的对曲线段进行校正处理, 再进行移动, 压缩, 拉伸, 剔除异常曲线段等操作.

3.1 分类校正

3.1.1 曲线段分类

根据人工校正特点, 人工判断会根据每段曲线段的尖峰数目, 曲线段最大值位置, 我们可以根据这两个特点对曲线段进行分类, 这两个特征是第一级特征, 本文分成四类. 第一类是尖峰数目为一个并且最大值点在曲线段的前半部; 第二类为尖峰数目一个并且最大值点在曲线段的后半部; 第三类为尖峰数目大于一个并且最大值点在曲线段前半部; 第四类为尖峰数目大于一个并且最大值点在曲线段后半部. 如图 2 的分

类图.

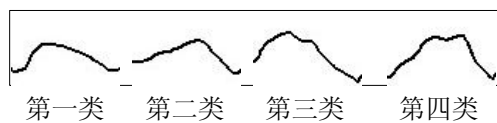


图 2 曲线段分类图

我们把最大值点在中间的分在前半部分, 如果一段曲线最大值点数目大于一个, 就把深度最小的最大值点设为这段曲线的最大值点. 本文用数字 1、2、3、4 分别代表第一类、第二类、第三类、第四类, 以便标记.

3.1.2 曲线段匹配

经过曲线段的分类后, 曲线就可以用一串数字表示, 比如图的一段 SSP 可以表示为“321142141”而 CPI 表示为“2421142141”, 下一步求两个曲线的公共子集, 比如 SSP 和 CPI 的公共子集为“21142141”. 设两个集合分别为 A 和 B. 计算机计算公共集的步骤为:

- ① 设集合 A 和 B, 取集合 A 的第一个元素 a_1 .
- ② 在集合 B 中从第一个元素 b_1 开始查找与 a_1 相同的数字, 依次标记为 b_i, b_j, \dots .
- ③ 取 A 中的第二个元素 a_2 , 在上面查找到的 b_i, b_j, \dots 的下一个元素开始查找, 使得 $a_1 a_2 = b_i b_{i+1}$.
- ④ 取 A 中的第三个元素 a_3 , 继续搜索, 直到 B 中没有元素和 A 匹配. 记录下搜寻后的最长集合 $a_1 a_2 \dots$ 和对应的 $b_i b_{i+1} \dots$.
- ⑤ 取 A 中的第二个元素 a_2 , 跟 a_1 一样在集合 B 里搜寻, 找到最长以 a_2 开头的最长子集 $a_2 a_3 \dots$ 和对应的 $b_j b_{j+1} \dots$.
- ⑥ 取 A 中的第三个元素 a_3 查找, 直到遍历 A 中所有元素. 找出一个或多个子集 $a_i a_{i+1} a_{i+2} \dots$ 和对应的 $b_j b_{j+1} b_{j+2} \dots$.

找到对应的公共子集后, 分类就完成了. 大部分曲线段就能对应上, 为后面校正的操作(移动, 拉伸, 压缩)做好了准备.

3.2 曲线段相似性

经过上面的分类, 两条曲线大部分曲线段会一一对应上, 然而可能会出现异常点或者大尺度分类有个别曲线匹配没有匹配上, 需要更精确的相似性匹配.

3.2.1 曲线特征提取

如图 1 所示, 比较的两条曲线不能以简单的斜率、厚度、峰值等来判断相似性, 本文将以曲线段的最大值上下厚度比为主要的特征, 因为, 不管曲线是移动、

伸长或者压缩了, 最大值上下厚度比是不变的. 本文还把相对斜率作为影响特征, 这里的最大值上下厚度比和相对斜率为第二级特征. 相对斜率的定义是根据测井解释的显示来的, 测井解释中, SSP 和 CPI 所测的幅度值不一样, 解释人员为了能够比较两条曲线会调整显示宽度, 使得两条曲线看上去幅度值差不多, 对阈值进行了调整, 如图 1 所示, SSP 的阈值是 $[-25, 25]$, CPI 阈值是 $[0, 40]$. 所以我们定义特征向量 $e_i = (\text{最大值上下厚度比}, \text{相对斜率})$.

最大值上下厚度比:

$$P = (X_{\max_i} - X_{i-1}) / (X_i - X_{\max_i})$$

相对斜率:

$$K = (CPI_{\max} - CPI_{\min} / SSP_{\max} - SSP_{\min}) * (X_{\max_i} - X_{i-1} / Y_{\max_i} - Y_{i-1})$$

$$e_i = (\alpha p, \beta k)$$

X_{\max_i} 为 CPI 一段曲线段的最大值点的深度值; X_{i-1} 为 CPI 一段曲线段的起始点深度值; X_i 为一段曲线段的终止点的深度值; CPI_{\min} 、 CPI_{\max} 为 CPI 曲线的左右刻度值, SSP_{\min} 、 SSP_{\max} 为曲线 SSP 的左右刻度值; Y_{\max_i} 为 CPI 一段曲线曲线段的最大值点的幅度值; Y_{i-1} 为 CPI 一段曲线曲线段的起始点的幅度值. α , β 分别为最大值上下厚度比和相对厚度的影响因素.

3.2.2 曲线相似性度量

曲线段的相似性可以通过距离度量, 欧拉距离是最常用的距离测量公式之一, 加权的欧拉距离的结果与所取的权系数密切相关, 而权系数的确定具有很大的随意性和不确定性.

马氏距离是由印度统计学家马哈拉诺比斯(P. C. Mahalanobis)提出的. 它是建立在总体样本之上的一种统计距离, 马氏距离的优点是它不受原始数据的单位影响^[1], 而本文所比较相似性的两条曲线测量值单位就不一样, 适合运用马氏距离.

$$D(e_i, e_j) = (e_i - e_j) \sum^{-1} (e_i - e_j)^T$$

其中 e_i 和 e_j 是进行对比的两个曲线段为目标曲线段特征矩阵的协方差矩阵.

4 实验

本文采用大庆油田测试解释评价中心多口中子寿命井的实际测量曲线进行校正. 依照基于重要点对曲线进行分段, 然后运用分级特征进行对比较正. 实验中设置参数如下: 曲线 SSP 设置分段常量 $R_{SSP} = 0.2$, CPI 设置分段常量 $R_{CPI} = 0.16$. 最大值上下厚度比和

相对厚度的影响因素分别为 $\alpha = 1$, $\beta = 0.78$.

以大庆油田采油一厂“中 611-斜 307”为例. 深度校正的前后情况如图 3 所示, 前半部分为校正前两条曲线, 后半部分为校正后的两条曲线, 可以看到, 用分级处理校正能够很好的利用测井曲线的顺序性, 找出对应曲线段, 并且, 能够找出异常点, 这样下一步移动、拉伸、压缩曲线段, 删除曲线段就很容易实现了.

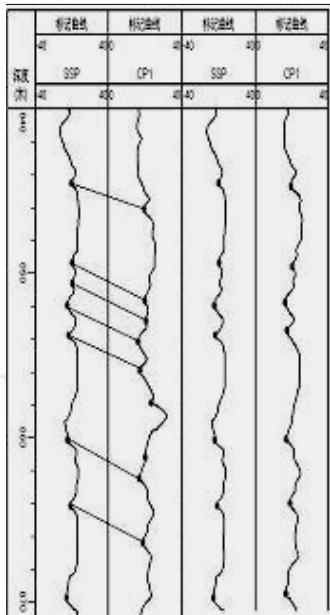


图 3 校正前后对比图

在校正过程中, 本实验将本文提出的分级对此校正的结果与人工校正的结果进行比较, 针对 10 口井的 322 个曲线段进行了对比较正, 校正结果如表 1 所示.

表 1 分级特征提取方法校正结果

井名	曲线段数目	正确率/%	分级方法校正时间	人工方法校正时间
中 611-斜 307	42	83.3	12s	>7min
拉 10-1566	42	78.6	11s	>6min
升 72-44	27	85.2	10s	>4min
南 1-丁 3-54	63	81.0	22s	>10min
南 214-平 287	12	100	10s	>3min
升 42-34	19	94.7	10s	>4min
北 1-331-61	60	85	22s	>10min
南 3-丁 1-斜	14	92.9	10s	>3min
南 3-丁 10-P457	18	100	11s	>3min
升 43-21	25	92	11s	>4min
综合	322	86.0	12.9s	>5.4min

实验结果表明, 本文提出的分级特征提取方法在

中子寿命测井曲线深度校正中正确率符合了解释标准, 而且较人工校正相比大大提高了校正效率.

5 结语

本文运用了测井曲线的顺序性的特点, 结合人工校正的经验, 合理地测井曲线进行分级校正, 解决了对比两条测量不同地层信息曲线提取特征困难的难题.

该方法对测井曲线段进行分类, 利用测井曲线具有顺序性的特点, 对分类没有对应的曲线段进行精度校正, 提高了算法运行的效率. 实验证明该方法能够有效的对曲线进行校正, 为测井资料进行下一步的解释做好了预处理.

该方法简单易懂、容易实现, 将该方法用于中子寿命测井的深度校正中, 本文校正正确率没有达到 100%, 校正后还得人工对校正错误曲线进行纠正. 实验表明该方法能较准确且快速校正曲线深度, 能大大提高中子寿命测井的深度校正的效率.

参考文献

- 1 郑希科, 韩玉堂, 袁纯予, 董秋燕, 李亚娟. 测-渗-测中子寿命测井在大庆油田的应用. 测井技术, 2004, S0(28): 48-52.
- 2 Suykens JAK, Gestel VT, Brabanter JD, et al. Least squares support vector machines. World Scientific. Singapore. 2002.
- 3 梁建海, 张建业. 基于斜率偏离的时间序列相似性搜索方法研究. 计算机应用研究, 2010, 27(1): 54-55.
- 4 尚福华, 马楠, 杜睿山. 基于形态特征的测井曲线相似性搜索研究. 计算机应用研究, 2013, 30(4): 1076-1081.
- 5 魏莲, 吴信才. 动态规划字符串匹配算法在曲线对比中的应用. 计算机工程与应用, 2007, 43(8): 8-9, 17.
- 6 封靖波, 苏志勋, 刘秀平. 一种基于相似曲线的全景图自动拼接算法. 计算机学报, 2003, 26(11): 1604-1608.
- 7 肖立志, 刘堂宴, 傅荣珊. 利用核磁共振测井评价储层的捕集能力. 石油学报, 2004, 25(4): 38-41.
- 8 胡魁贤, 严宏志, 朱自冰, 杨振容, 蒋燕英. 截面轮廓曲线分段约束拟合. 计算机工程与科学, 2010, 6(6): 153-155.
- 9 李国和, 刘文胜. 基于概率统计局部聚类的曲线分段方法. 计算机工程与应用, 2006, 13: 162-164, 171.
- 10 曾海泉. 时间序列挖掘与相似性查找技术研究[博士学位论文]. 上海: 复旦大学, 2003: 23-25.
- 11 Mao Y, Zhou XB, Xia Z, et al. A survey for study of feature selection algorithms. Pattern Recognition and Artificial Intelligence, 2007, 20(2): 211-218.