doi:10.3772/j.issn.1002-0470.2020.12.002

基于多粒度聚类的测井曲线自动分层识别方法①

姬庆庆②[*,**] 朱登明[**] 石 敏[***] 王兆其[**] 周 军[****]

([*]中国科学院大学 北京 100049)

([**]中国科学院计算技术研究所前瞻研究实验室 北京 100190)

([***]华北电力大学控制与计算机工程学院 北京 102206)

([****]中国石油集团测井有限公司 西安 710065)

摘 要 随着测井技术及大数据分析技术的快速发展,自动测井解释技术可以有效辅助人工快速开展储层划分、油水层解释等工作。为了提升储层划分及油水层识别准确度,本文提出了一种基于有监督学习的多粒度聚类识别方法,该方法通过对标准测井曲线及分层结果的学习提取不同分层测井曲线特征,在划分出储层的基础上再进行油水层识别。与已有方法相比,本文方法通过对真实测井曲线进行多种处理,从而融合曲线多层次特征,有利于取得更加准确的分层结果。实验结果表明,该方法可以对测井曲线进行自动分层,提高了曲线自动分层的效率,在真实测井曲线上能够取得较好的分层识别结果。

关键词 自动分层;多粒度聚类;测井曲线

0 引 言

测井技术在勘探过程中收集关于地质储集层的“四性”信息,即“岩性、物性、电性、含油性”,测井解释则需要通过“四性”之间的关系建立测井解释模型,确定油层有效厚度。测井解释通过研究储集层电性与岩性、物性、含油性的对应关系,力求消除岩石矿物背景对于油层信息的影响,从而达到客观评价砂岩储集性能和流体性质,并准确划分储层的目的[1]。

现阶段国内众多单位多以常规曲线为基础,结合区块地质特点,采用传统方法利用人工对油气层进行识别、解释和评价,还没有较为成熟的自动分层解释软件[2]。由于人工解释的方法需要大量人力,同时分层结果易受解释人员主观因素的影响,因此难以满足不断提高的测井解释要求。

随着油气勘探规模的不断扩大,测井解释将面对更多更复杂的研究对象,测井分层解释方法的发展与测井技术的发展息息相关。利用计算机技术对测井曲线进行自动分层[3],可以在一定程度上避免专家依赖、减少人力消耗,是未来技术发展趋势。现阶段用于岩性自动识别的方法主要有概率统计法、聚类分析法、支持向量机法(support vector machine,SVM)[4-5]。概率统计法往往以概率统计为基础,通过先验和条件概率估计后验概率对测井进行识别解释。概率统计方法适用于物性特征条件较好的数字测井资料,针对岩心资料较少但测井资料较多的情况能够取得一定的效果[6],但这一方法存在着先验概率获得困难、人为影响因素大等缺点。

本文在充分了解测井曲线数据特点的基础上,提出了一种基于有监督学习的多粒度聚类算法的测井曲线自动分层识别方法。利用该方法可以对测井曲线进行“储层”和“非储层”自动识别,还能够在

① 国家重大科技专项(2017ZX05019005)资助项目。

② 男,1994 年生,博士生;研究方向:虚拟现实,大数据技术;联系人,E-mail:1602584488@qq.com

(收稿日期:2020-01-10)

“储层”内对测井曲线进行油水层自动识别。本文方法优势如下。现有测井曲线分层识别方法往往需要先对数据进行归一化操作[7-9]，而本文方法为了更好地发掘曲线特征，无需对曲线数据进行归一化操作，降低了程序运算的时间复杂度；现有方法往往只利用曲线本身特征开展分层识别，本文方法在充分分析测井曲线和油藏之间关系的基础上，引入关键属性导数、属性比值，通过这些内部特征发掘，有效提升了分层识别准确率；现有分层识别方法较少进行多粒度分层识别，而本方法能够先划分“储层”和“非储层”，然后在“储层”内进行细化识别。

聚类算法能够快速准确地解决各种分类问题，近年来受到了学者们的广泛关注。江超等人[10]利用 K-means 聚类方法对激光雷达采集的障碍物信息进行聚类，并结合其他方法实现机器人的实时避障功能。刘畅等人[11]将聚类算法应用于监控视频运动目标检测与感兴趣区域分割问题的研究。这些研究往往将聚类算法作为某一步骤进行使用，缺少针对具体数据进行深入分析并对聚类算法进行改进的探索。为解决测井曲线自动识别问题，本文深入分析测井数据特点，并结合问题本身对聚类算法进行改进，从而能够在测井曲线分层识别问题中获得更好的表现。

1　测井数据预处理

1.1　测井曲线数据

丰富和准确的测井曲线是进行准确自动分层的基础，随着近些年测井技术的不断发展，随钻测井技术方法和仪器在不断创新和完善。与此同时，在测井过程中，还逐步引入密度、电成像等技术，勘探开发和技术人员能够收集到更加全面的随钻测井资料并形成测井曲线。随钻测井技术所形成的测井曲线既能够为勘探工作提供导向，还能够根据地质属性评价油气资源情况，已经成为石油勘探公司关注和研究的热点技术[12-14]。上述测井技术的发展使得获取的测井曲线数据更加真实，更有利于开展自动分层识别工作。

图 1 展示的是青海某区块的测井数据及解释结论，图中对测井技术在勘探过程中收集的关于地质储集层的“岩性、物性、电性、含油性”进行了展示。其中，左侧 3 道为测井过程中能够体现“岩性、物性、电性”的原始测井曲线，右侧第 1 道为“含油性”信息。在人工测井曲线标定过程中，往往通过观察测井曲线特点并结合标定人员经验对测井曲线进行分层标定，而自动测井解释正是需要利用计算机技术在它们之间建立联系。

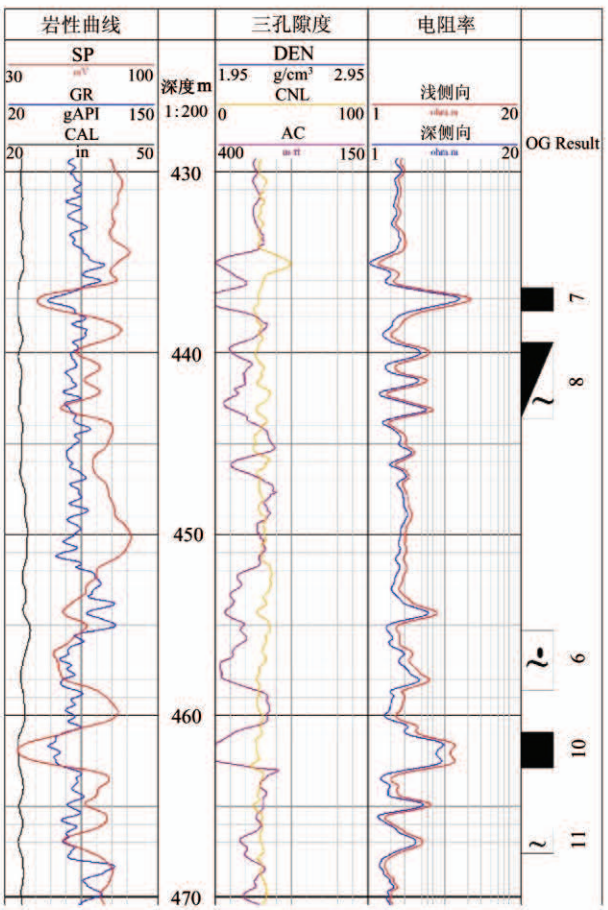

图 1　测井曲线数据示意图

1.2　测井曲线关键属性遴选

主成分分析算法是通过求解目标协方差矩阵的前 N 个最大特征值对应的特征向量来组成特征映射矩阵从而实现样本的主成分空间映射[15-16]。伴随着测井设备的发展，测井设备能够采集到越来越丰富的测井数据。测井数据涉及到多个物理属性，但并非所有的属性都与地质分层具有相关性，因此需要对复杂多样的测井数据与地层分布的相关性进行判断。利用主成分分析法能够对测井数据与地质

分层的相关性做出很好的判断,经过主成分分析法的分析,再引入专家系统辅以修正,能够选择出具有重要影响的测井属性。同时主成分分析也能够起到对高维测井数据降维、有效减少运算复杂度的作用。本文方法在数据关键信息缺失最少的原则下,探寻原始测井数据中对于测井分层结果具有重要影响的测井属性,将这些重要指标称为主成分。

针对测井数据而言,某测井样本包含 m 个测井属性,即 m 个维度,每个测井属性沿深度方向有 n 个采样点,这样便构成了 $m \times n$ 阶矩阵,如此大的数据规模计算机处理起来存在一定的困难。因此需要在这些繁杂的数据信息中分析出能够表征地质含油特性的有用信息,也就是需要在 m 维数据中寻找关键属性数据。为了有效解决这一问题,需要对测井数据进行降维操作,把原来较多测井曲线反映出的信息进行简化,然后使用少数几个相互独立互不相关的综合指标代替,同时还需要有限的几个指标能够尽量充分反映原来多指标信息。

设 $\boldsymbol{\Sigma}$ 为根据测井属性建立的特征矩阵,通过运算 $\boldsymbol{\Sigma}$ 可得的特征值为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant 0$, $\boldsymbol{e}_1 = (e_{i1}, e_{i2}, \cdots, e_{im})^{\mathrm{T}}$ 为对应特征值 λ_i 的正交单位特征向量,则第 i 个主成分的贡献率为

$$C_{\lambda i} = \frac{\lambda_i}{\sum_{k=1}^{m} \lambda_i} \tag{1}$$

累计贡献率为

$$\sum C_{\lambda i} = \frac{\sum_{k=1}^{i} \lambda_i}{\sum_{k=1}^{m} \lambda_i} \tag{2}$$

测井区块所涉及的测井数据包含 60 个不同的测井属性维度,本文按照累计贡献率 >85%、特征值大于 1 的原则选取主成分,选择前 7 个主成分进行分析。在主成分分析法中,主成分载荷矩阵指的是各项原始指标与主成分之间的关系,某一原始指标与主成分联系越紧密,则经过荷载矩阵计算出的联系系数的绝对值越大。因此主成分荷载矩阵可以反映出测井曲线分层的主要影响因素。

本文首先对原始测井曲线进行主成分分析,然后再通过主成分荷载矩阵分析各原始测井曲线与主成分之间的联系,进而选取用于测井曲线自动分层的测井曲线。经过上述选择,最终确定用于测井曲线自动分层的测井曲线为声波(AC)、自然电位(SP)、中子(CNL)、密度(DEN)、自然伽马(GR)、深侧向电阻率(LLD)和浅侧向电阻率(LLS)等 7 条测井曲线。

1.3 测井属性数据预整理

受测井设备及地质构造的限制,测井初始段和结束段往往会出现数值异常的测量值,例如99 999、-99 999、0 等。这些异常值会对测井曲线自动分层识别带来负向影响,因此在分层之前就需要对这些异常数据做剔除处理。

在测井过程中,由于使用多种测井设备,因此存在如下 3 种采样深度间隔:0.075 m、0.1 m 和0.125 m。在同一深度下能够更好地对每个采样点的多种属性特征进行提取,本文利用克里金插值算法,对不同采样间隔的曲线统一采样间隔达到深度校正的目的,从而保障在相同深度下,不同测井属性都存在相应的属性值,方便对层位特征信息提取分析。经过深度校正与统一,本文采样点深度间隔为 0.1 m。

对于储层划分及测井解释这一问题而言,根据人工解释的经验,往往在重要曲线出现大幅波动的深度段,会有层位点出现。根据前述证据权重法分析可知,自然电位曲线在储层及非储层划分中能够起到重要的作用,因此本文引入自然电位曲线的一阶导数、二阶导数参与测井曲线“储层”与“非储层”划分。

构建由自然电位曲线一阶导数、二阶导数构成的向量:

$$\vec{\boldsymbol{D}} = \{\dot{x}_{SP}, \ddot{x}_{SP}\} \tag{3}$$

由于曲线异常数据点已经在前期处理过程中剔除,因此有效避免了曲线不可导的情况。

在一些地质条件下,对于水层受泥浆侵入影响不同的情况,测井曲线深侧向电阻率(HLLD)比浅侧向电阻率(HLLS)的值升高更多,即二者比值(HLLD/ HLLS)应大于油层的二者比值。故可以根据淡水泥浆侵入对深侧向电阻率和浅侧向电阻率影响的不同,利用二者比值深入发掘测井曲线特征从而提升储层划分及油水层解释准确率[17]。

2　测井曲线自动识别

2.1　测井曲线自动分层识别模型

针对测井曲线自动分层识别问题，本文采用图 2所示技术路线开展研究。

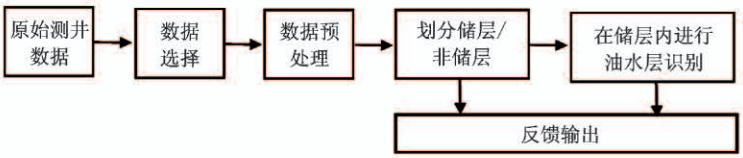

图 2　测井曲线自动分层识别技术路线

K-均值聚类算法是聚类分析中常用的方法，其核心思想是在 n 个样本的数据库中，给定分类数 K，然后选取 K 个初始聚类中心，计算样本点到各个聚类中心的欧氏距离，并将样本点划分到距离它最近的聚类中心所属的一类中去。当各个样本点所属类别都划分完毕后，利用计算平均值的方式得到每一类样本点新的聚类中心，通过不断迭代，直至两次迭代得到的聚类中心相同，此时获得最终的聚类中心并完成分类。

对于测井曲线自动分层问题，将测井曲线根据不同曲线特征划分为不同储层，其本质上是一种聚类问题。本文结合测井数据具体特点，采用 K-均值聚类算法对测井数据进行分类。在测井曲线自动分层问题中，其基本思路是给定储层分类和各个测井属性代表值（即聚类中心），计算测井数据到聚类中心的欧式距离，然后将其归类于距离最小的聚类中心对应的储层类别。

在测井曲线自动分层识别前，先对测井数据进行数据处理，这一步骤包括剔除无效数据、测井数据深度对齐、重点属性一阶求导、二阶求导等步骤。由于数据采集过程存在多种因素干扰，从而使得某些数据出现异常波动，因此还需对数据作滤波操作，去除测井数据中的异常值。

传统的无监督聚类算法只需要给出样本数据即可通过不断寻找聚类中心完成聚类，但对于测井数据而言，由于各测井属性之间相似性极高，利用无监督的聚类算法很难准确找到合适的聚类中心，同时也会带来运算时间代价的提升[18-20]。针对上述问题，本文将有监督的机器学习思想与 K-聚类分析法相结合，构建 K-最近邻（K-nearest neighbor，KNN）网络。因为本文所涉及的测井数据中存在人为标定好的准确分层信息，这些人为标定好的分层信息可以作为聚类学习过程中训练数据的数据标签。在有标签标定的训练数据引导网络寻找最佳聚类中心的基础上，能够更好地引导 K-均值聚类问题分别寻找到不同标签类别对应的聚类中心，从而完成对各不同标签类别的识别任务，该方法流程图如图 3 所示。

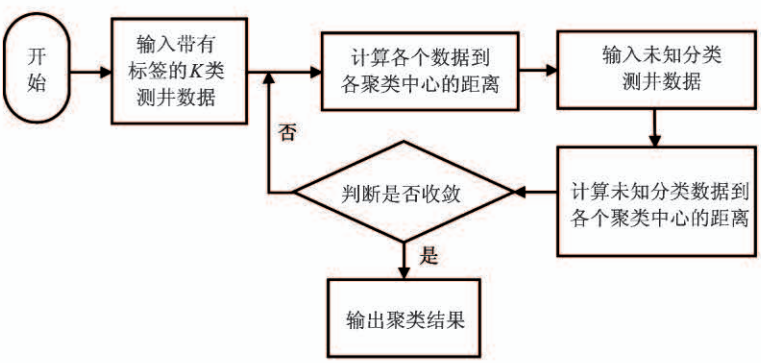

图 3　KNN 方法流程图

在本方法中，首先输入经过预处理的带有人工标定标签的 K 类测井数据，然后利用聚类算法计算各个数据样本点到各个标签类别对应的聚类中心的距离。通过不断迭代找出各个标签类别对应的最佳

聚类中心,从而完成有监督的聚类算法的训练过程。接下来对待分类样本点进行分类标定,首先输入待分类样本点对应的测井数据,然后利用聚类算法计算各待分类样本点到各个类别聚类中心的距离,最终将待分类样本点划归至欧氏距离最小的聚类中心对应的类别。待所有待分类样本点全部识别标定结束后,输出聚类结果。

2.2 储层划分

本文在进行测井曲线分层识别过程中,采用2.1节提出的KNN算法,并将多粒度识别的思想与之结合。

首先构造向量:

$$\vec{\boldsymbol{Y}}_1 = \{x_{AC}, x_{SP}, x_{GR}, \vec{\boldsymbol{D}}\} \tag{4}$$

由于测井曲线之间的差异能够较为准确地划分"储层"及"非储层",因此为了更加快速地对测井曲线进行"储层"和"非储层"划分,本文先对 $\vec{\boldsymbol{Y}}_1$ 利用有监督的聚类算法对测井曲线进行"储层"与"非储层"划分,得出相应结果。

中值滤波是一种非线性的滤波技术,相比于线性滤波器,它在一定程度上能够去除数据的局部异常[21-22]。对数字信号序列 $X_j(-\infty < j < \infty)$ 进行滤波处理时,先要定义长度为奇数的 n 长窗口,$n = 2N + 1$,N 为正整数。设在某一时刻,窗口内的信号样本为 $x(i-N),\cdots,x(i),\cdots,x(i+N)$,其中 $x(i)$ 为位于窗口中心的信号样本值。对这 L 个信号样本值按从小到大的顺序排列后,其中在 i 处的样本值定义为中值滤波的输出值。

对于本文测井曲线"储层"及"非储层"识别而言,其识别结果为"0"、"1"两种数值结果,其中"0"代表"非储层","1"代表"储层"。结合中值滤波具有在滤除数据噪声的同时能够较好地保护连续数据数值的特点,本文选用中值滤波对测井曲线"储层"及"非储层"识别结果进行滤波处理,从而达到对个别预测误差点进行修正的目的。

2.3 储层内油水层识别

由于"油水层"在地质构造、地层物性等方面有着众多相似点,因此需要引入更多测井曲线挖掘其特征从而对"油水层"进行识别。

为了实现"油水层"识别,构造向量:

$$\vec{\boldsymbol{Y}}_2 = \{x_{CNL}, x_{DEN}, x_{LLD}, x_{LLS}, x_{\frac{LLD}{LLS}}\} \tag{5}$$

由于"油水层"仅出现在"储层"段内,因此将"非储层"数据剔除,仅保留"储层"数据段,这样一方面能够加快"油水层"自动识别速率,另一方面能够排除"非储层"数据带来的负向影响,从而提升"油水层"自动识别精度。在"储层"数据段内,对 $\vec{\boldsymbol{Y}}_1$ 和 $\vec{\boldsymbol{Y}}_2$ 构成的数据集合利用前述有监督学习的聚类算法进行"油水层"识别,即利用已有人工标定结果的样本点测井属性数据,寻找不同层位的聚类中心。然后将待分类样本点对应的测井属性与各聚类中心比较,计算欧氏距离,从而将待分类样本点划分为干层、水层、油层、含油水层、差油层和油水同层等6类层位。

这种做法能够排除具有大量数据点位的"非储层"数据的干扰,同时最大限度地放大"储层"内油水层之间的数据特征,能够使多粒度聚类算法取得更好的"油水层"识别结果。多粒度聚类算法完成"油水层"识别后,再次使用滤波算法对少量误差点进行滤波操作,纠正识别过程中部分零星点位的误差,然后输出最终的"油水层"识别结果。

3 实验设计及结果分析

本文选用青海地区某区块内10口垂直井测井数据作为实验数据,井位分布如图4所示,其中9口井的"储层"、"非储层"及"油水层"划分数据作为训练数据,对另外一口井即Y189井的"储层"和"非

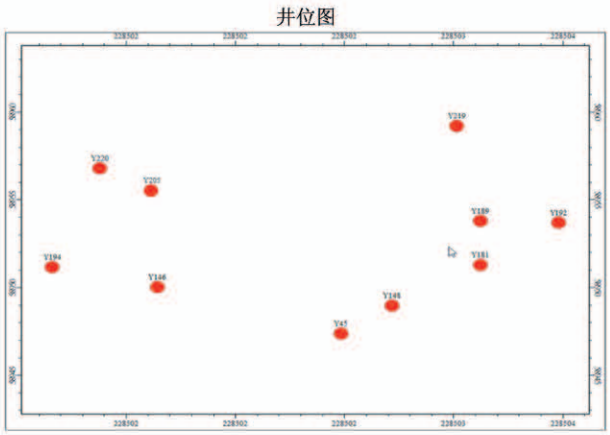

图4 实验用井井位图

储层”两类层位分布进行识别，然后在“储层”数据段内对“油水层”进行识别。

3.1　“储层”与“非储层”识别

实验前，选取 Y220、Y219、Y205、Y194、Y192、Y181、Y148、Y146、Y45 等 9 口测井的声波、自然电位、自然伽马 3 条原始测井曲线及自然电位曲线的一阶导数、二阶导数进行实验，不同井的曲线深度范围不完全相同，但大都集中在 500 ~ 1 300 m 范围内，采样间隔为 0.1 m。预测井 Y189 井的预测范围为 500 ~ 1 100 m，总长 600 m，共需预测 6 000 个采样点。

KNN 聚类分析法针对 Y189 井“储层”和“非储层”识别结果图如图 5 所示，本文以670.3 ~ 744.8 m、1 041.5 ~ 1 095.2 m 深度段的预测结果为例进行说明。图 5 中，左侧 3 道为本文方法所使用的测井曲线随深度变化在不同地层产生不同的响应值；右侧第 1 道为滤波前的“储层”、“非储层”预测结果，右侧第 2 道为滤波后的“储层”、“非储层”预测结果，图中有颜色填充的深度段属于“储层”，未进行填充的深度段属于“非储层”。右侧第 3 道为人工标注的识别结果，有符号标注的深度段为“储层”，未进行标注的深度段为“非储层”。

如表 1 所示，表格中展示了人工标定的分层结果、本文方法自动识别的分层结果及二者之间的误差值，由该表数据可以看出，在大部分层位本文方法能够很好地识别“储层”与“非储层”。

3.2　多粒度油水层识别

在剔除“非储层”数据段后，本方法继续进行油水层划分。该实验区域储层内共有干层、差油层、水层、含油水层、油层、油水同层等 6 类层位。实验过程中，将 Y220、Y219、Y205、Y194、Y192、Y181、Y148、Y146、Y45 等 9 口测井的声波、自然电位、中子、密度、自然伽马、深侧向电阻率和浅侧向电阻率等 7 条原始测井曲线及深侧向电阻率/浅侧向电阻率作为训练数据输入。表 2 是多粒度油水层识别结果。

Y189 井包含干层、水层、油水同层和油层等 4 类层位。对于 Y189 井的油水层识别，从表 2 中可以看出，尽管本文多粒度聚类识别方法针对“油层”、

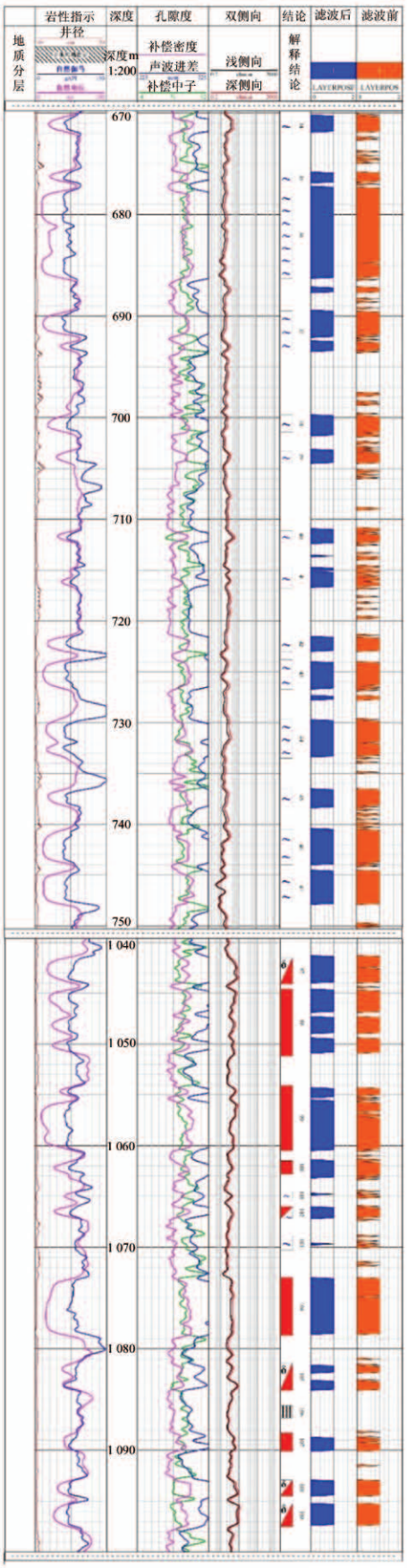

图 5　测井曲线“储层”、“非储层”识别结果

表 1　Y189 井“储层”自动识别结果

人工标注储层深度段/m		本文方法识别储层深度段/m		实验误差/m	
起始深度	结束深度	起始深度	结束深度	上界误差	下界误差
…					
670.3	672.1	670.2	671.9	0.1	0.2
675.8	677.0	675.8	677.0	0.0	0.0
677.7	686.2	677.2	686.2	0.5	0.0
689.5	693.5	689.5	693.5	0.0	0.0
699.6	701.4	699.6	701.8	0.0	0.4
703.2	704.3	703.1	704.5	0.1	0.2
711.1	712.2	711.0	712.4	0.1	0.2
714.6	716.8	714.6	716.8	0.0	0.0
721.5	723.0	721.5	723.0	0.0	0.0
723.8	726.8	723.9	726.8	0.1	0.0
729.7	733.5	729.7	733.5	0.0	0.0
736.5	738.3	736.5	738.3	0.0	0.0
740.5	744.0	740.5	744.0	0.0	0.0
744.8	747.8	744.5	747.8	0.3	0.0
…					
1041.5	1044.1	1041.2	1044.1	0.3	0.0
1044.6	1051.2	1044.6	1049.9	0.0	0.3
1054.1	1060.5	1054.2	1060.5	0.1	0.0
1061.5	1062.8	1061.5	1063.2	0.0	0.4
1066.0	1067.2	1066.0	1067.2	0.0	0.0
1073.0	1078.6	1073.0	1078.6	0.0	0.0
1081.4	1084.1	1081.8	1084.2	0.4	0.1
1088.4	1090.1	1088.8	1090.1	0.4	0.0
1092.8	1094.6	1092.8	1094.6	0.0	0.0
1095.2	1097.5	1095.2	1097.5	0.0	0.0
…					

表 2　Y189 井多粒度油水层识别结果

储层深度段/m		真实类别	本文方法识别结果
起始深度	结束深度		
…			
670.2	671.9	水层	水层
675.8	677.0	水层	水层
677.2	686.2	水层	水层
689.5	693.5	水层	水层
699.6	701.8	水层	水层
703.1	704.5	水层	水层
711.0	712.4	水层	水层
714.6	716.8	水层	水层
721.5	723.0	水层	水层
723.9	726.8	水层	水层
729.7	733.5	水层	水层
736.5	738.3	水层	水层
740.5	744.0	水层	水层
744.5	747.8	水层	水层
…			
1 041.2	1 044.1	油层	油层
1 044.6	1 049.9	油层	油层
1 054.2	1 060.5	油层	油水同层
1 061.5	1 063.2	油水同层	水层
1 066.0	1 067.2	油层	油层
1 073.0	1 078.6	油层	油层
1 081.8	1 084.1	油层	油水同层
1 088.8	1 090.1	油层	油层
1 092.8	1 094.6	油层	油层
…			
1 280.4	1 281.7	干层	干层
1 288.1	1 290.4	干层	干层
1 296.5	1 298.8	干层	油层
1 309.0	1 312.5	干层	干层
1 324.5	1 330.3	干层	油层
1 335.5	1 337.1	干层	干层
…			

“干层”这两类层位识别时存在一定程度的误差,但在 29 个层位当中,本文设计的多粒度聚类识别方法能够准确识别 24 个层位,识别准确率已经达到 82.8%。说明利用多粒度聚类识别方法将“非储层”数据去除,然后在“储层”内继续进行识别的方法能够获得更加准确的识别效果。

3.3　对比实验

在分类问题领域中,卷积神经网络(convolutional neural networks,CNN)算法常被用于解决图像分类问题,支持向量机(SVM)算法往往被用于解决文本分类问题。为体现本文所选方法在解决测井曲线自动分层问题上的优势,将本文方法与常见的SVM 算法、经典 CNN 算法在同一真实测井曲线数据集中开展对比实验。其中 SVM 算法与传统 CNN 算法采用青海某区块 100 口井数据进行训练,然后对 Y189 井进行储层划分及油水层识别,实验结果如表 3 所示。此处仅选取 670.3 ~ 712.2 m、1 041.5 ~ 1 094.6 m 深度段进行展示。

通过对比可以发现,本文方法相较于 SVM 算法在“储层”边界划分上精确度更高;二者在水层的识别上表现相当,但在其他层位的识别中,本文方法取

表 3 Y189 井不同算法对比实验结果

人工标注储层深度段/m		本文方法识别储层深度段/m		SVM 算法识别储层深度段/m		CNN 算法识别储层深度段/m		真实类别	本文方法识别油水层结果	SVM 算法识别油水层结果	CNN 算法识别油水层结果
起始深度	结束深度	起始深度	结束深度	起始深度	结束深度	起始深度	结束深度				
						.....					
670.3	672.1	670.2	671.9	669.8	672.3	669.9	672.5	水层	水层	水层	水层
675.8	677.0	675.8	677.0	675.4	677.2	675.6	677.4	水层	水层	水层	水层
677.7	686.2	677.2	686.2	677.5	686.9	677.4	686.5	水层	水层	水层	水层
689.5	693.5	689.5	693.5	689.1	693.7	689.4	694.2	水层	水层	水层	水层
699.6	701.4	699.6	701.8	699.3	702.3	699.1	702.2	水层	水层	水层	水层
703.2	704.3	703.1	704.5	702.7	704.8	702.9	704.5	水层	水层	水层	水层
711.1	712.2	711.0	712.4	710.6	712.7	710.4	712.8	水层	水层	水层	水层
						.....					
1041.5	1044.1	1041.2	1044.1	1041.0	1044.5	1040.6	1044.2	油层	油层	油层	油层
1044.6	1051.2	1044.6	1049.9	1044.3	1050.4	1044.0	1050.2	油层	油层	油层	油水同层
1054.1	1060.5	1054.2	1060.5	1053.8	1060.7	1054.0	1061.2	油层	油水同层	油水同层	油水同层
1061.5	1062.8	1061.5	1063.2	1061.2	1063.6	1061.2	1063.5	油水同层	水层	水层	油水同层
1066.0	1067.2	1066.0	1067.2	1065.4	1067.5	1065.3	1068.4	油层	油层	油水同层	油水同层
1073.0	1078.6	1073.0	1078.6	1073.2	1079.4	1072.9	1079.8	油层	油水同层	油水同层	油水同层
1081.4	1084.1	1081.8	1084.2	1082.0	1084.7	1082.5	1084.9	油层	油水同层	水层	水层
1088.4	1090.1	1088.8	1090.1	1087.9	1090.8	1087.3	1091.2	油层	油层	油层	油层
1092.8	1094.6	1092.8	1094.6	1092.5	1095.2	1092.3	1095.2	油层	油层	油水同层	油水同层
						.....					
1280.4	1281.7	1280.2	1281.8	1280.0	1281.9	1280.1	1281.9	干层	干层	干层	干层
1288.1	1290.4	1288.1	1290.3	1287.9	1290.5	1288.1	1290.4	干层	干层	干层	油层
1296.5	1298.8	1296.4	1298.8	1296.3	1299.0	1296.2	1299.1	干层	油层	油层	油层
1309.0	1312.5	1309.1	1312.5	1309.2	1312.5	1308.8	1312.8	干层	干层	油层	干层
1324.5	1330.3	1324.5	1330.3	1324.3	1330.4	1324.3	1330.4	干层	油层	油层	油层
1335.5	1337.1	1335.5	1337.2	1335.4	1337.2	1335.6	1337.4	干层	干层	干层	干层
						.....					

得了更加准确的识别结果。同时本文方法相较于经典 CNN 算法在上界划分中表现相近,但在下界划分中取得了较好的结果;在油水层识别中,经典 CNN 算法对于油层及干层的识别不够敏感,误差较大,识别效果劣于本文方法。上述对比实验说明本文方法在实验测井数据上能够比 SVM 算法及经典 CNN 算法取得更好的实验结果。

为了更好地衡量本文方法在测井曲线“油水层”划分问题中的时间消耗,将本文方法与 CNN 算法及 SVM 算法进行比较。其中本文方法选取 9 口井为训练集,CNN 算法及 SVM 算法选取 100 口井作为训练集,在训练结束后均选择同一口井在 3 种方法中对训练结果进行测试。实验结果如表 4 所示。

表 4 不同算法时间消耗对比实验结果

	本文方法	SVM 算法	CNN 算法
模型训练用时/s	4 201	5 949	5 474
测试用时/s	5	4	4

通过上述实验可以发现,SVM 算法所需训练及测试时间最长,其次是 CNN 算法,所需时间最少的为本文方法。尽管本文在训练阶段采用 9 口测井数

据进行训练,但在测井曲线划分精度优于其他两种算法的同时,时间消耗上也优于其他两种算法,说明本文方法具有较强的先进性和实用性。

4 结 论

本文提出了一种基于聚类算法的测井曲线自动分层识别方法。在对测井数据进行剔除异常值、统一采样间隔、求取重要属性一阶导数和二阶导数等预处理的基础上,通过多粒度聚类的方法将测井曲线先自动划分为"储层"、"非储层",然后在"储层"内对测井曲线开展进一步的识别。该方法相对于传统方法具有以下优势。第一,本方法无需对测井曲线进行归一化操作,能够降低算法运行时间复杂度;第二,本方法引入重要属性导数、基于测井属性的比值数据等,能够更加充分地挖掘测井曲线特征,取得更好的分层识别结果;第三,该方法在有监督的聚类算法基础上能够对测井曲线实现多粒度识别,在提高测井曲线油水层识别准确率的基础上,还能够满足实际生产中的不同需求。

本文所提出的方法已在真实测井数据上开展实验并取得了较好的自动分层识别结果,但目前对于油水混合类层位的识别准确度还不够高。未来将针对这些问题进行改善,从而使本文方法能够更加精准地识别油水混合类层位,同时在石油测井解释领域取得更加广阔的应用。

参考文献

[1] 张海涛. 苏里格地区有效储层测井识别方法研究[D]. 西安:西北大学地质学系,2010

[2] Goldberg D, Meltser A. High vertical resolution spectral gamma ray logging: a new tool development and field test results[C]//SPWLA 42nd Annual Logging Symposium, Houston, USA, 2001: 1-13

[3] 王楠. 测井曲线模式识别及其在地层对比中的应用[D]. 哈尔滨:黑龙江大学数学科学学院, 2008

[4] 钟广法,马在田. 测井序列变点分析自动识别岩性界面[J]. 天然气工业, 2006(1):46-48,160

[5] 彭涓. 基于极差分析法在测井曲线自动分层问题中的研究[J]. 自动化与仪器仪表, 2015(1):36-38

[6] 李保霖. 基于测井数据的岩性识别方法研究[D]. 西安:西安科技大学理学院,2012

[7] Cui M M, Fan A P, Wang Z X, et al. A volumetric model for evaluating tight sandstone gas reserves in the permian Sulige gas field, Ordos basin, central China[J]. *Acta Geologica Sinica (English Edition)*, 2019, 93(7):386-399

[8] Liu Z, Wang Z, Liu J, et al. Logging response characteristics and reservoir significance of volcanic rocks in the eastern sag of Liaohe depression[J]. *Journal of Jilin University*, 2018, 48(1):285-297

[9] Wang Y M, Li X J, Chen B, et al. Lower limit of thermal maturity for the carbonization of organic matter in marine shale and its exploration risk[J]. *Petroleum Exploration and Development*, 2018, 45(3):28-37

[10] 江超, 邢科新, 林叶贵, 等. 未知环境下移动机器人静态与动态实时避障方法研究[J]. 高技术通讯, 2019,29(10):1012-1020

[11] 刘畅, 王鹏钧, 赵潇, 等. 监控视频 ROI 完整分割技术研究[J]. 高技术通讯, 2019,29(11):1053-1062

[12] Ghiselin D. Service providers advance MWD/LWD technologies[J]. *Offshore*, 2014,74(8):60-62

[13] Thorsen A K, Eiane T, Thern H F, et al. Magnetic resonance in chalk horizontal well logged with LWD[J]. *SPE Reservoir Evaluation and Engineering*, 2010, 13 (4): 654-666

[14] Wang H, Fehler M. The wavefield of acoustic logging in a cased hole with a single casing—Part II: a dipole tool [J]. *Geophysical Journal International*, 2018, 212(2): 1412-1428

[15] Jiang J, Wen Z, Zhao M X, et al. Series arc detection and complex load recognition based on principal component analysis and support vector machine[J]. *IEEE Access*, 2019, 7: 47221-47229

[16] Hapsari Y. Weather classification based on hybrid cloud image using principal component analysis (PCA) and linear discriminant analysis (LDA)[J]. *Journal of Physics Conference Series*, 2019, 1167:012064

[17] 秦绪英. 电阻率测井泥浆侵入影响校正方法研究——以鄂尔多斯盆地天然气储层为例[J]. 石油物探, 2006(5):537-541,18

[18] Lu X S, Zhou M C, Qi L, et al. Clustering-algorithm-based rare-event evolution analysis via social media data

[J]. *IEEE Transactions on Computational Social Systems*, 2019, 6(2):1-10

[19] 周冰钰,刘博,王丹,等. 基于自组织中心 K-means 算法的用户互动用电行为聚类分析[J]. 电力建设, 2019,40(1):68-76

[20] Chen F Y, Yang Y C, Xu L W, et al. Big-data clustering: K-means or K-indicators? [J]. *arXiv*: 1906.00938, 2019

[21] Amanipour V, Ghaemmaghami S. Median filtering forensics in compressed video[J]. *IEEE Signal Processing Letters*, 2019, 26(2):287-291

[22] Green O. Efficient scalable median filtering using histogram-based operations[J]. *IEEE Transactions on Image Processing*, 2018, 27(5):2217-2228

Automatic layering recognition method for logging curves based on multi-granularity clustering

Ji Qingqing[*,**], Zhu Dengming[**], Shi Min[***], Wang Zhaoqi[**], Zhou Jun[****]

([*]University of Chinese Academy of Sciences, Beijing 100049)

([**]Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

([***]College of Control and Computer Engineering, North China Electric Power University, Beijing 102206)

([****]China Petroleum Logging Co., Ltd., Xi'an 710065)

Abstract

With the rapid development of logging technology and big data analysis technology, automatic logging interpretation technology can effectively auxiliary artificial rapid reservoir classification and oil-water reservoir interpretation. In order to improve the accuracy of reservoir classification and oil-water layer identification, this paper gives a method based on multi-granularity of supervised learning. This method extracts the characteristics of different layers of logging curves by learning the standard logging curves and the layered results, and then identifies the oil and water layers based on the reservoir division. Compared with the existing methods, the proposed method can fully explore the characteristics of the real logging curve by processing the real logging curve through a variety of treatments, which is conducive to obtain more accurate stratification results. The experimental results show that the proposed method can automatically separate the logging curves, improve the efficiency of the automatic slicing of the curves, and can obtain good results of the layer identification on the real logging curves.

Key words: automatic layering, multi-granularity clustering, logging curve