

# 利用决策树方法识别复杂碳酸盐岩岩性

——以苏里格气田苏东41-33区块为例

王振洲<sup>1</sup>, 张春雷<sup>2</sup>, 高世臣<sup>1</sup>

(1.中国地质大学(北京)数理学院,北京 100083; 2.北京中地润德石油科技有限公司,北京 100083)

**摘要:** 苏里格气田苏东41-33区块具有低孔、低渗透、非均质性强的特点,其奥陶系马家沟组碳酸盐岩储层受多期不同类型构造、沉积等作用,岩性复杂多样,对岩性的准确识别成为研究区开发的重点和难点。近年来,决策树方法作为一种机器学习方法,在地质学领域中的运用越来越受到关注,尤其在岩性预测方面。以测井、录井资料为基础,通过岩性参数特征分析,优选出对岩性较敏感的声波时差、自然伽马、光电吸收截面指数、密度、深侧向电阻率和补偿中子6种测井参数,通过分析这6种测井参数特征,构建基于决策树方法的多分类器,将岩性信息与岩石特征信息融合。与测井、录井资料的对比、分析结果表明,利用决策树方法识别复杂碳酸盐岩岩性的准确率超过80%,且相比朴素贝叶斯方法,其岩性识别的准确率提高了13%。

**关键词:** 复杂碳酸盐岩 测井岩性识别 决策树方法 机器学习方法 苏里格气田

中图分类号: TE112.2

文献标识码: A

文章编号: 1009-9603(2017)06-0025-09

## Lithology identification of complex carbonate rocks based on decision tree method: An example from Block Sudong41-33 in Sulige Gas Field

Wang Zhenzhou<sup>1</sup>, Zhang Chunlei<sup>2</sup>, Gao Shichen<sup>1</sup>

(1.School of Science, China University of Geosciences(Beijing), Beijing City, 100083, China;

2.Beijing Zhongdi Runde Petroleum Technology Co., Ltd., Beijing City, 100083, China)

**Abstract:** The reservoir of Block Sudong41-33 in Sulige Gas Field has the characteristics of low porosity, low permeability and high heterogeneity. The carbonate reservoir in Lower Ordovician Majiagou Formation were subjected to multi-stage multi-type construction, sedimentary and other effects, which makes the lithology complex and diverse, and thus the accurate identification of lithology has become a difficult problem of development in this area. In recent years, more and more attention has been focused on the use of decision tree method in machine learning in the field of geoscience, especially in lithology prediction. Based on the data of well logging and the analysis of lithological parameters, six kinds of well logging parameters that are sensitive to lithology were selected, which includes acoustic time difference (AC), natural gamma ray (GR), photoelectric absorption cross section index (PE), density (DEN), deep lateral resistivity (RLLD) and compensated neutron (CNL). Through the analysis of the six well logging parameters, a multi classifier was constructed based on decision tree method, and the information of lithology and rock characteristics were fused. Compared with the lithologic data of well logging, the recognition accuracy is over 80%. When compared with the Naive Bayesian, the accuracy of lithology recognition is improved by 13%.

**Key words:** complex carbonate rock; lithology identification of well logging; decision tree method; machine learning method; Sulige Gas Field

岩性识别是地层评价、油藏精细描述以及储层地质建模的重要研究内容之一,钻井取心是岩性识

收稿日期:2017-07-11。

作者简介:王振洲(1993—),男,山东鄄城人,在读硕士研究生,从事数据挖掘及地球物理勘探研究。联系电话:15901036656, E-mail: gush i1004@sina.com。

别最直接有效的方法,但其成本较高,且对每口井取心不符合实际。因此,基于测井资料利用间接方法开展岩性特征研究成为岩性识别的重要手段,其方法主要涉及机器学习、人工智能、模式识别、统计学等多个研究领域,主要有神经网络、支持向量机、贝叶斯网络等方法<sup>[1-3]</sup>。其中,神经网络方法已广泛应用于基于测井资料的岩性识别,但神经网络收敛速度慢、准确率较低,易对训练样本过度拟合<sup>[4]</sup>,且其可变参数太多,以至于将训练集本身的一些特点当作所有数据都具有的一般性质,因此该方法没有实际应用价值。支持向量机方法为二分类算法,当其解决多分类问题时需利用其他方法进行扩展,且支持向量机方法的分类效果还受其参数和核函数选取的影响。贝叶斯网络方法是基于条件独立性提出的,但往往与输入变量相关,从而导致准确率较低。此外,神经网络、支持向量机和贝叶斯网络等方法均属于“黑箱”模型,其样本的数据和属性之间仅具有笼统的因果关系,但对其内部规律却是未知的<sup>[5]</sup>。

神经网络、支持向量机和贝叶斯网络等方法在对复杂碳酸盐岩的岩性识别过程中均存在一定的缺陷,而决策树方法对复杂碳酸盐岩岩性识别则具有显著优势。决策树方法属于“白箱”模型,其对表述的要素或过程的规律具有清楚的认识,对于各相关因素也具有较为深刻的表征,对实际岩性预测具有很好的指导作用。为了验证决策树方法的可行性,以苏里格气田苏东41-33区块为例,以测井、录井资料为基础,通过岩性参数特征分析,优选出对复杂碳酸盐岩岩性较敏感的声波时差、自然伽马、光电吸收截面指数、密度、深侧向电阻率和补偿中子6种测井参数,使用决策树方法对研究区奥陶系马家沟组五段(马五段)复杂碳酸盐岩的岩性进行识别。

## 1 决策树方法

### 1.1 基本原理

决策树方法是一种树状结构的分类预测模型<sup>[6]</sup>,在有标记的训练集中学习分类规则,以预测待识别目标的类型为目。从根节点开始,每个内部节点(非叶子节点)表示在一种属性上对训练数据进行测试,每一个分支代表该测试的一个输出。根据不同的测试结果将训练集划分为不同的训练子集,每个训练子集构成一个叶子节点,而叶子节点存放一个分类标号,代表最终的分类结果。决策树

的生成是一个递归划分训练集的过程,其划分训练集的原则是使杂乱无序的数据集变得有序,其中衡量数据集无序程度常用的指标有信息熵和基尼指数。

信息熵是度量数据样本集无序程度最常用的一项指标。假定当前数据样本集  $D$  中第  $k$  类数据样本所占的比例为  $p_k$ ,则  $D$  的信息熵定义为

$$Ent(D) = -\sum_{k=1}^n p_k \log_2 p_k \quad k=1, 2, \dots, n \quad (1)$$

式中:  $Ent(D)$  为数据样本集  $D$  的信息熵;  $D$  为数据样本集编号;  $k$  为不同类型的数据样本编号;  $n$  为数据样本集个数;  $p_k$  为第  $k$  类数据样本占数据样本集的比例。

分类回归树算法是决策树方法中最常用的算法之一,其根据基尼指数选择划分属性,进而对数据样本集进行划分<sup>[7]</sup>。数据样本集的无序程度可以根据基尼指数来度量,其计算式为

$$Gini(D) = \sum_{k=1}^n \sum_{k' \neq k} p_k p_{k'} \quad (2)$$

式中:  $k'$  为不同类型的数据样本编号;  $Gini(D)$  为数据样本集  $D$  的基尼指数。

一般而言,数据样本集的信息熵越小,则其无序程度越高,划分所得的结果也越准确。而基尼指数反映的是从数据样本集中随机抽取2个样本,其类别标记不一致的概率。因此,基尼指数越小,数据样本集的无序程度越高。

分类回归树是一种在数据挖掘领域应用较多的二叉决策树。该算法常用于对数据进行分类预测,且对于连续型和离散型的目标变量均适用。通过该算法对数据样本的训练,最终生成一个分类决策树,根据该决策树的分支条件,可以为无标签样本集的有效分类预测提供依据。

数据样本划分是分类回归树生成过程中较重要的步骤。目前应用较为广泛的度量方法是以基尼指数为依据的无序程度度量方法,针对每个数据样本集属性划分的目的是确定最小基尼指数,基尼指数越小,其属性无序程度越高,进而可以得到最优的属性划分结果。

### 1.2 岩性识别模型建立

为了反映利用决策树方法识别复杂碳酸盐岩岩性的效果,首先将数据样本集分为训练集和测试集,通过对训练集的数据样本进行训练,根据其计算得到的信息熵和基尼指数来确定决策树的分支节点,进而形成决策树,并通过决策树的各分支节点和叶子节点组成分类器,最终对测试集中所有的

数据样本进行岩性预测。利用决策树方法建立岩性识别模型的具体流程包括根据训练集生成决策树和根据决策树识别岩性2部分。其中,根据训练集生成决策树的具体流程为:①创建根节点,为分支节点分配类别,遍历训练集1次。②如果训练集均属于同一类别或训练集中只剩下1个样本,则返回分支节点,并将其作为叶子节点,为其分配属性。③对训练集循环执行每个属性划分,计算该属性划分的基尼指数,并比较各个属性的基尼指数,取基尼指数最小的属性作为分支节点,将训练集划分为2个子集。④对训练集划分的2个子集重复步骤①—③,直至基尼指数为0,即该分支节点全为一类,标记为叶子节点,输出分类器参数结果。根据决策树识别岩性的具体流程为:①遍历测试集,并储存至数据中。②获取决策树以及用于构建决策树的标签向量。③比较测试数据与决策树上的数值,递归执行该过程直至进入叶子节点。④将测试数据定义为叶子节点所属的类型。⑤依次合并分类结果,输出最终的分类结果和决策树模型。

## 2 应用实例分析

### 2.1 区域概况

苏里格气田苏东41-33区块下古生界马五段属于华北海型沉积,主要发育碳酸盐岩储层,气藏前期生产特征显示,储层的含气性和产能主要受控于岩性及物性。分析测井、录井资料发现,研究区岩石类型主要为白云岩、灰岩、泥岩及其之间的过渡岩性,可进一步划分为灰岩、白云质灰岩、泥质灰岩、白云岩、灰质白云岩、泥质白云岩和泥岩7种类型,含气储层主要分布于灰质白云岩、白云质灰岩等岩性。由于研究区碳酸盐岩储层岩性成分的复杂性和各向异性,给岩性解释带来很大困难,因此针对复杂碳酸盐岩的岩性识别是研究区储层评价的关键。

### 2.2 测井参数优选

测井参数是对岩石多种物理性质变化的综合反应<sup>[8-10]</sup>,对于不同的地层,其对应的测井参数的差异主要取决于岩性(组成岩石的矿物成分)、岩石的颗粒大小以及孔隙中的流体性质。反之,一组特定的测井参数也对应于地层中的某一种或几种岩性。测井参数包含丰富的岩性信息,不同的测井参数对岩性具有不同的区分度<sup>[11-14]</sup>,其中对岩性反映灵敏的有声波时差、自然伽马、光电吸收截面指数、密度、深侧向电阻率和补偿中子6种测井参数。

测井参数数据统计结果及其对各类岩性的敏感性分析结果(图1)表明,不同岩性类型具有不同的测井参数响应特征。随着泥质含量的增加,声波时差和自然伽马增大,而深侧向电阻率则减小。泥岩的自然伽马多大于80 API,与其他类型的岩性具有较大差异,能够明显区分,因此为了更好地展示其他岩性与测井参数的响应关系,泥岩在图1中没有显示。泥质灰岩和泥质白云岩的自然伽马也相对较高,多为30~80 API,但二者的界限不是特别明显。光电吸收截面指数和深侧向电阻率对灰岩的反映相对较为敏感,而其他4种测井参数则对灰岩的反映很不敏感。白云岩的光电吸收截面指数主要为2.5~3.5 b/e,明显低于灰岩;而补偿中子则相反,灰岩的补偿中子明显小于白云岩。对于灰质白云岩和白云质灰岩则难以根据单一测井参数进行划分,因此单一测井参数的线性划分不能对复杂碳酸盐岩岩性进行分类,亟需综合多种测井参数进行岩性的识别。

通过主成分聚类等方法<sup>[15-18]</sup>得到的岩性分类结果不易于从地质学角度进行分析。决策树方法能够高度自动化地建立易于地质解释人员理解的岩性识别模型,且其模型属于“白盒”封存,地质解释人员可以直接利用专业知识进行测井参数-岩性的相关分析,有助于理解模型并发现新的信息和规律;在样本空间中多维度上利用决策树方法划分局部子空间,进而在子空间中进行岩性识别。根据理论和实际分析结果,选取6种测井参数建立针对研究区复杂碳酸盐岩岩性的测井参数样本集。

### 2.3 测井参数概率密度分布特征

根据不同岩性的测井参数概率密度分布特征,可以得出不同岩性测井参数的大致分布范围。对于灰岩和白云岩可以通过测井参数概率密度分布特征来区分,而灰质白云岩和白云质灰岩、泥质灰岩和泥质白云岩仅通过单一测井参数概率密度分布特征却不易区分。为了更加清晰地表征不同岩性的测井参数概率密度分布特征,在6种测井参数中任选2种,构成散点图矩阵;对该散点图矩阵进一步分析得到二维中心坐标以及协方差矩阵,然后得出二维测井参数90%置信度的置信椭圆(图2)。置信椭圆可以去除2种测井参数之间的异常点,将大部分具有代表性的测井参数显示出来;且置信椭圆宽度越窄,表明2种测井参数之间的相关性越好。例如图2中泥质灰岩和泥质白云岩的声波时差和自然伽马的置信椭圆宽度较窄,呈较强的正相关性;通过这种较好的正相关性即可明显地区分含泥质

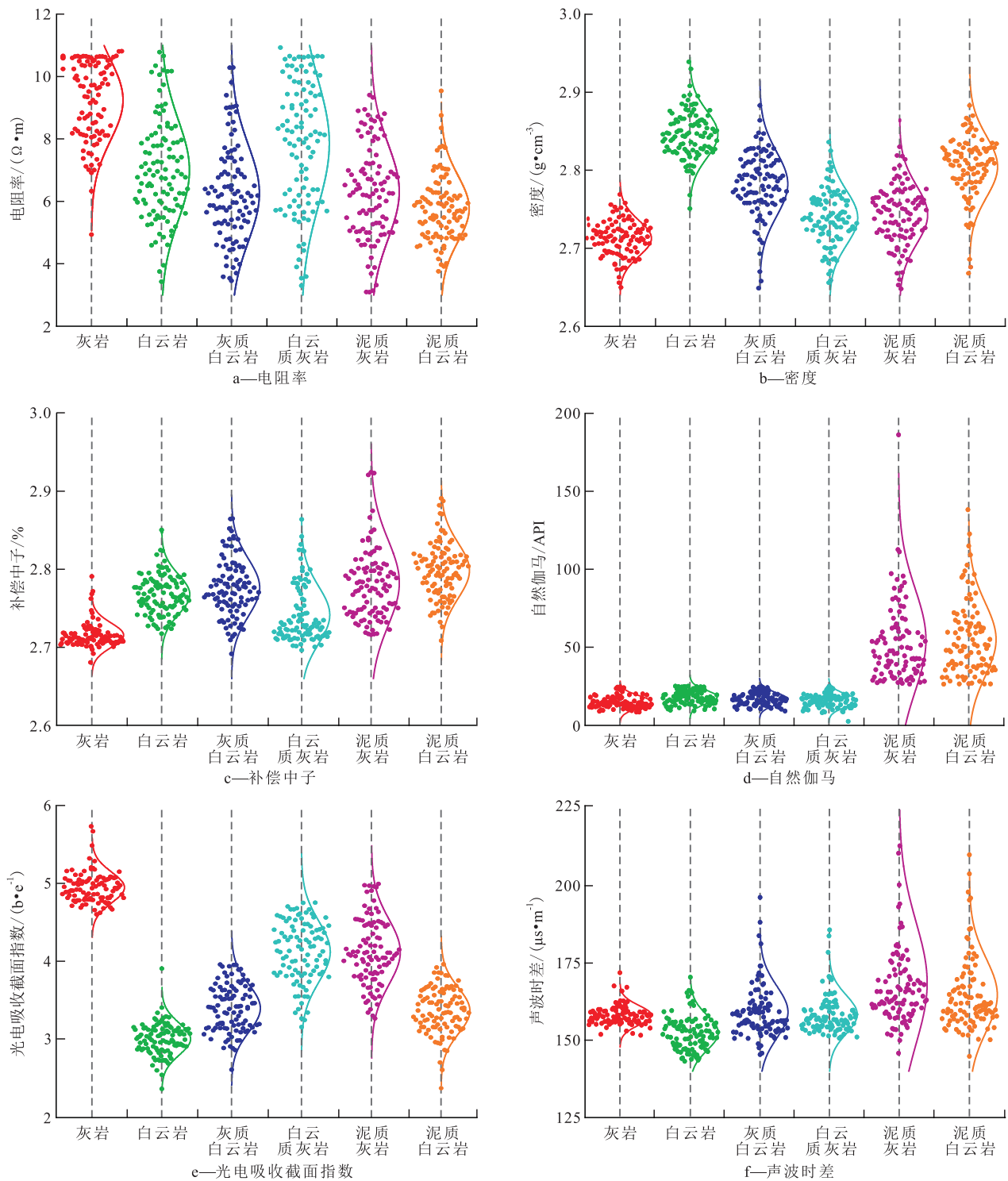


图1 测井参数的岩性敏感性分析

Fig.1 Sensitivity analysis of well logging parameters to lithology

岩性,且与图1的岩性敏感性分析结果一致。分析补偿中子和深侧向电阻率90%置信度的置信椭圆发现,各种岩性置信椭圆的宽度均较窄,呈较强的负相关性,且灰岩和白云岩置信椭圆的重合面积较小,因此根据补偿中子和深侧向电阻率的相关性即可以区分灰岩和白云岩,该结果也进一步验证了测井参数岩性敏感性分析结果的准确性。

### 2.4 岩性识别模型建立与分析

为了反映决策树方法识别岩性的效果,将苏东41-33区块所有测井参数数据汇总,建立测井参数样本集,针对各类岩性随机抽取100个测井参数样本,形成训练集。综合分析测井参数概率密度分布特征,选择声波时差(*AC*)、自然伽马(*GR*)、光电吸收截面指数(*PE*)、密度(*DEN*)、深侧向电阻率(*RT*)和

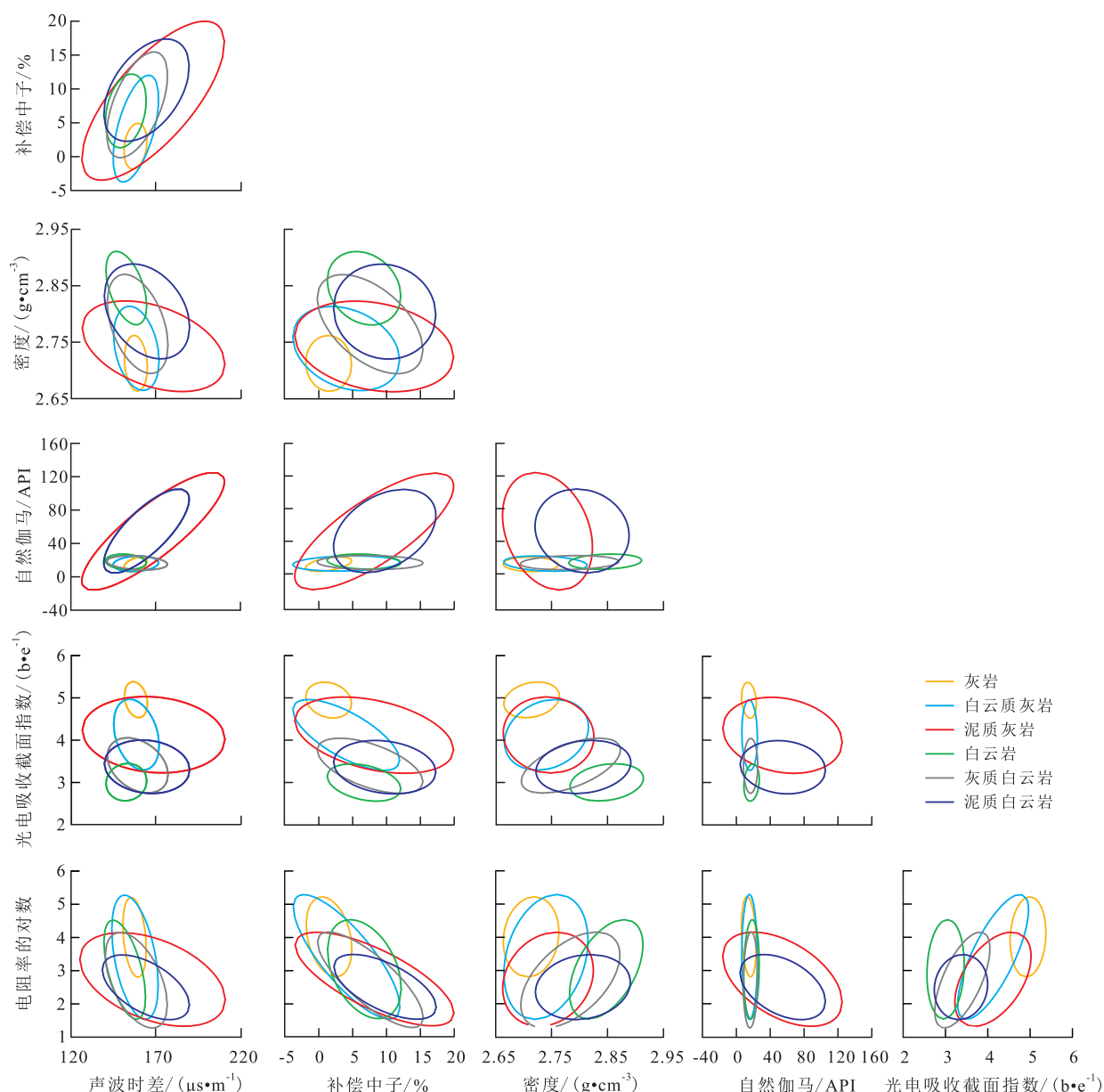


图2 二维测井参数90%置信度的置信椭圆

Fig.2 Two-dimensional cross plots of well logging parameters (confidence ellipse with 90% confidence regions)

补偿中子(CNL)6种测井参数进行岩性分类、识别。最后对筛选的测试数据集,依据决策树方法构建复杂碳酸盐岩的岩性分类器,并将研究区中53-09井作为测试集进行岩性识别。

通过建立的测井参数岩性识别模型对训练集进行划分,可以得到基于决策树方法的岩性分类器(图3)。在利用决策树进行岩性分类的过程中,测井参数样本集的信息熵不断变小,测井参数样本的岩性类别划分逐渐清晰。通过决策树方法得到的测井参数岩性识别模型较为直观,不仅可以获得清晰的岩性类别划分,还可得到明确的物理属性划分及各种岩性的测井参数概率密度分布特征,为地质解释人员根据测井资料进行岩性识别提供依据。

通过对训练集的划分,得到基于决策树方法的复杂碳酸盐岩岩性识别模型。利用该模型对岩性进行分割,使得各子集各岩性类型的信息熵更小。首先选择信息熵变化最大的自然伽马进行分割,以自然伽马为24.96 API作为节点,将整个训练集分为2个子集;训练子集1为非泥质岩性(图3左侧分支),自然伽马小于或等于24.96 API,训练子集2为含泥质岩性(图3右侧分支),自然伽马大于24.96 API。由训练子集2以及随之得到的岩性划分结果(图4)可以看出,基于岩性识别模型的进一步划分,使岩性分类更加清晰。其中,当密度小于等于2.67 g/cm<sup>3</sup>时为泥岩(为更好地表征复杂碳酸盐岩岩性与测井参数的响应关系,该部分泥岩在图4a中未显

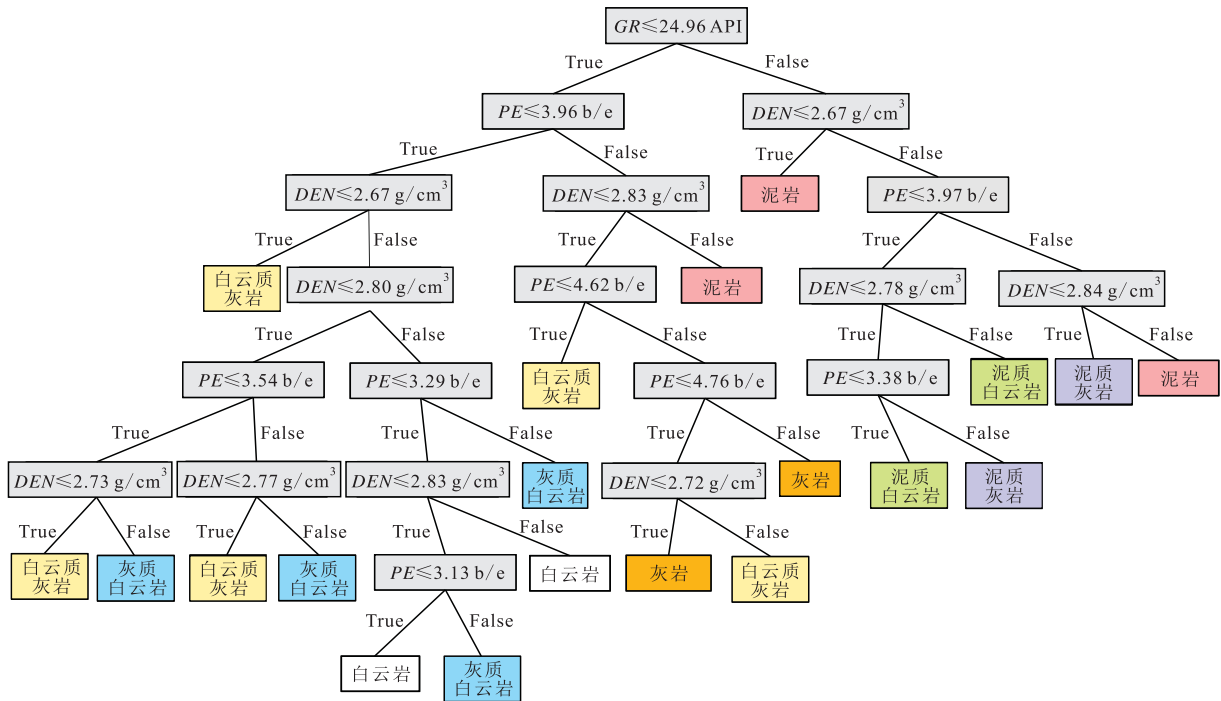


图3 基于决策树方法的岩性分类器

Fig.3 Lithology classifier based on decision tree method

示),而当密度大于  $2.67 \text{ g/cm}^3$  时,由于密度测井参数的信息熵变化较小,因此选择信息熵变化最大的光电吸收截面指数对训练子集2进行划分。当光电吸收截面指数大于  $3.97 \text{ b/e}$ ,密度为  $2.67 \sim 2.78 \text{ g/cm}^3$ ,为泥质灰岩的概率较大;密度为  $2.78 \sim 2.84 \text{ g/cm}^3$ ,为泥质白云岩的概率较大;密度大于  $2.84 \text{ g/cm}^3$ ,为泥岩的概率较大(图4a)。以光电吸收截面指数为  $3.38 \text{ b/e}$  作为节点对训练子集2进行岩性划分,结果(图4b)表明,当光电吸收截面指数大于  $3.38 \text{ b/e}$  时,为泥质灰岩的概率较大;而当光电吸收截面指数小于等于  $3.38 \text{ b/e}$  时,为泥质白云岩的概率较大。因此,根据每个决策树分支,可以将测井数据划分为

若干个子集,由多个决策树分支构成基于决策树方法的复杂碳酸盐岩岩性识别模型,可以更加清晰地表征岩性分类的结构特征,在一定程度上提高了复杂碳酸盐岩岩性识别的精度。

变量子空间是指在多维数据空间中任意变量组合构成的空间。对于多维数据空间,其子空间较多,所包含的信息较为广泛。将声波时差、自然伽马、光电吸收截面指数、密度、深侧向电阻率和补偿中子构成的六维全变量集分为多个子集,依据信息熵,作为划分准则构建决策树分类器。通过分类器得到的岩性分类结果显示,基于决策树方法的岩性识别模型对于不同测井参数子集的岩性分类精度

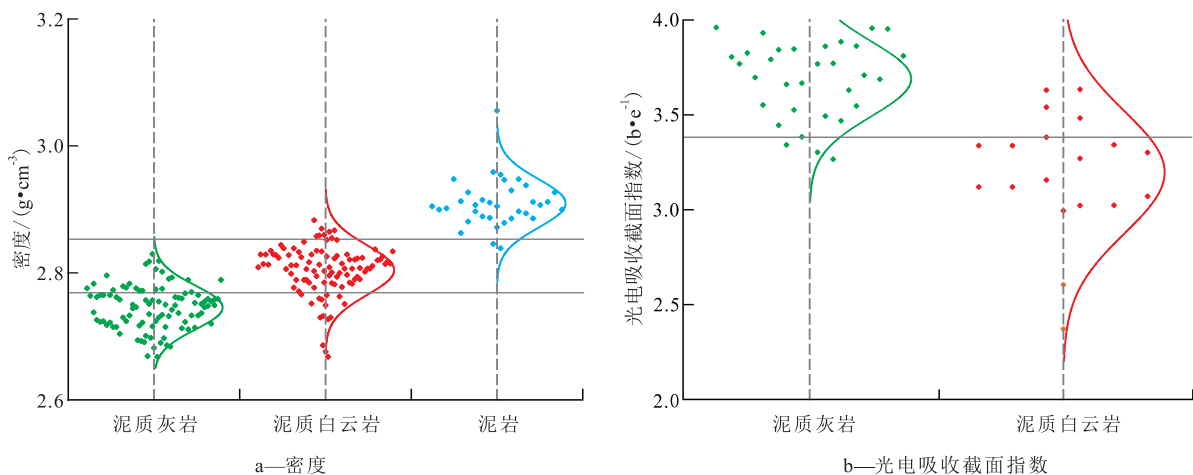


图4 基于岩性识别模型的岩性划分结果

Fig.4 Lithologic classification results based on lithology identification model

存在差异。在6种测井参数组成的多维全变量集中,基于信息论度量信息的指标划分杂乱无章的数据,将无序数据变为有序数据。从训练集中提取研究区测井参数对岩性约束的基本规律,结合地质解释人员的经验、认知,根据测井参数信息可以更加准确地识别岩性,也为地质解释人员发现新的规律和认识提供了新的思路。

### 2.5 岩性识别结果

将决策树方法与朴素贝叶斯方法对复杂碳酸

盐岩岩性的识别结果进行对比、分析。朴素贝叶斯方法是基于条件独立性提出的,可以通过特征之间的条件独立性假设,降低对数据量的需求,其输入的数据变量往往是相关的,但朴素贝叶斯方法仍然是一种有效的岩性分类方法。

以苏里格气田苏东41-33区块53-09井为例。图5为利用决策树方法和朴素贝叶斯方法对53-09井的岩性识别结果,将其与测井、录井分析的岩性结果对比发现,对于厚度较小的薄层,例如埋深为

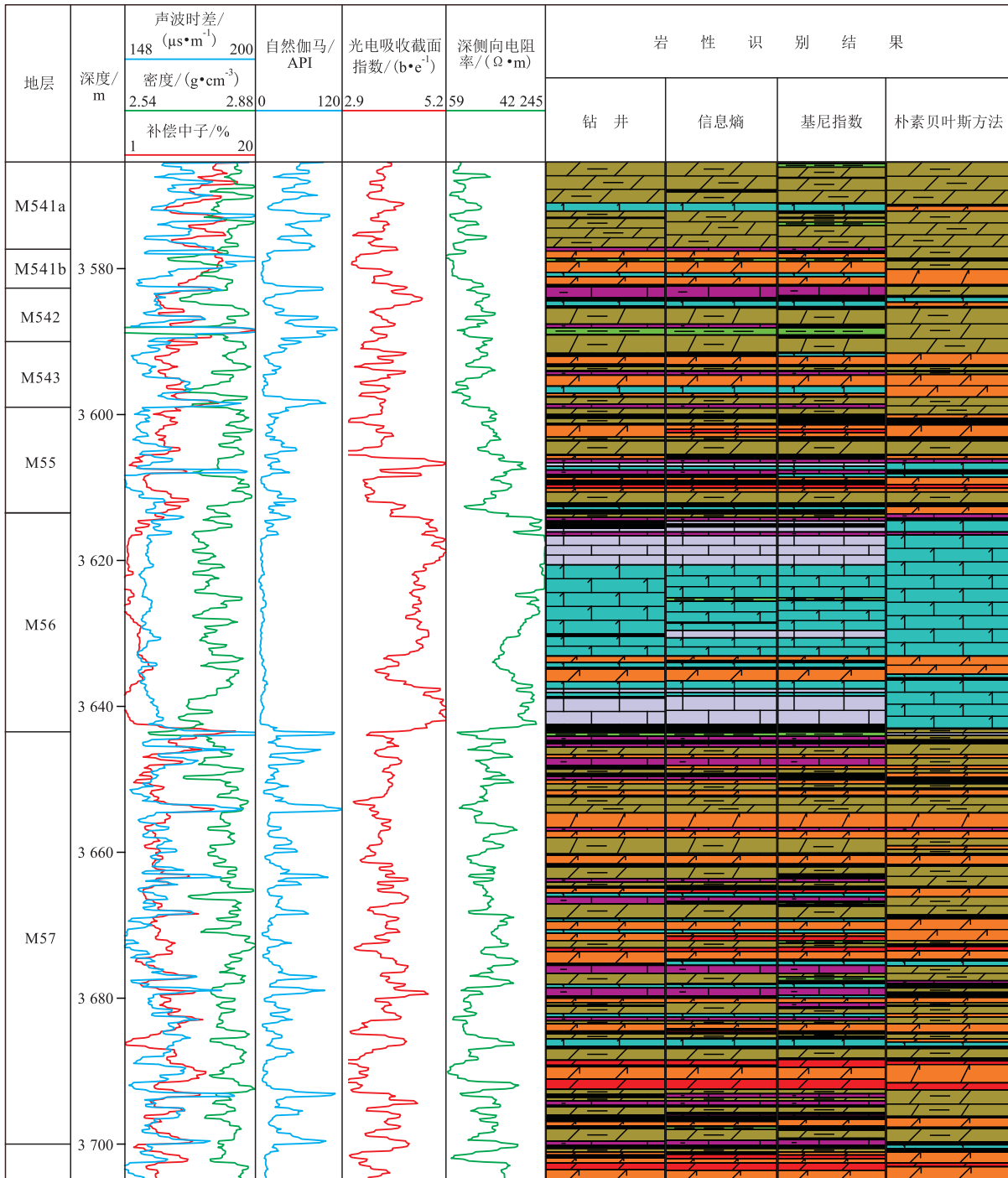


图5 苏里格气田苏东41-33区块53-09井岩性识别结果对比

Fig.5 Comparison chart of lithologic interpretation results in Well 53-09 of Block Sudong 41-33 in Sulige Gas Field

3 588~3 589 m处的泥质灰岩,利用朴素贝叶斯方法将其误判为灰岩,而利用决策树方法的识别结果与测井、录井的岩性解释结论一致。利用决策树方法预测的岩性序列特征与钻井岩性序列特征一致,主要为泥岩—泥质白云岩—白云质灰岩—白云岩。

分析决策树方法和朴素贝叶斯方法的岩性识别结果,得到53-09井不同岩性的混淆概率统计结果(表1),发现利用决策树方法建立的复杂碳酸盐岩岩性识别模型对该井岩性识别的准确率达93.12%,而朴素贝叶斯方法识别的准确率仅为66.41%。朴素贝叶斯方法对灰岩和泥质灰岩的预测效果较差,有66.12%的泥质灰岩被误判为泥质白云岩。而决策树方法对含泥质岩性的识别效果较

好,对泥质灰岩和泥质白云岩的正判率分别为96.20%和94.20%。53-09井白云岩类含量较多,其中泥质白云岩和灰质白云岩占总岩性的比例分别为32.70%和22.59%;而泥岩和泥质灰岩的含量较少,分别为3.64%和3.17%。对于这种不均衡数据,朴素贝叶斯方法表现出一定的局限性,其对含量较高的泥质白云岩和灰质白云岩的正判率分别为87.40%和89.55%,而对含量较低的泥岩和泥质灰岩的正判率仅为63.83%和15.85%。而决策树方法对这4类岩性的正判率均超过89.60%。利用决策树方法对研究区复杂碳酸盐岩岩性识别的效果整体较好,其准确率超过80%,且相比朴素贝叶斯方法,准确率提高了13%。

表1 苏里格气田苏东41-33区块53-09井不同岩性识别方法混淆概率统计结果  
Table1 Statics of confusing probability using different lithology identification methods in Well53-09 of Block Sudong41-33 in Sulige Gas Field

录井岩性	朴素贝叶斯方法预测岩性判别比例/%							决策树方法预测岩性判别比例/%						
	灰岩	白云质灰岩	泥质灰岩	白云岩	灰质白云岩	泥质白云岩	泥岩	灰岩	白云质灰岩	泥质灰岩	白云岩	灰质白云岩	泥质白云岩	泥岩
灰岩	0	100	0	0	0	0	0	100.00	0	0	0	0	0	0
白云质灰岩	0	63.04	0	0	36.23	0.73	0	5.40	89.90	0	0	2.50	0	2.20
泥质灰岩	1.09	7.65	15.85	0	5.46	66.12	3.83	0	0	96.20	0	0	0.50	3.30
白云岩	0	0	0	50.88	49.12	0	0	0	0	0	94.70	5.30	0	0
灰质白云岩	0	4.78	0	0	89.55	5.67	0	0	2.40	0	7.50	89.50	0	0.60
泥质白云岩	0.41	0.21	2.68	1.45	4.75	87.40	3.10	0	0	4.80	0	0	94.20	1.00
泥岩	12.77	0	0	0	0	23.40	63.83	0	0	0	0	0	0	100.00

### 3 结论

决策树分类器的建立不需要任何地质知识或参数设置,因此适合于探测式知识发现。决策树方法可以处理多维数据,获取的决策树模型直观且易于被人理解。决策树方法归纳的学习和分类步骤是简单和快速的。一般而言,决策树分类器具有很高的准确率。实验结果表明,决策树方法的应用有利于复杂碳酸盐岩岩性的预测,主要表现在以下几个方面:①在复杂碳酸盐岩岩性识别过程中,利用决策树方法得到的结果更加符合地质沉积环境,与传统岩性识别方法相比,获取的信息更加丰富,有利于地质解释人员发现新的信息和规律。②通过对岩性参数特征的分析,明确不同岩性的测井参数特征,进而优选出对岩性比较敏感的声波时差、自然伽马、光电吸收截面指数、密度、深侧向电阻率和补偿中子6种测井参数。③决策树方法具有鲁棒性和信息量大的特点,应用该方法建立的岩性识别模

型对研究区所有井进行岩性预测,得到的复杂碳酸盐岩岩性识别结果的准确率均超过80%。④利用决策树方法可以对复杂碳酸盐岩岩性识别达到定性的认识,但其分类目标是基于离散化的岩性类别,不能达到定量的描述,比如泥质含量、灰质含量。因此,须进一步研究通过连续目标的回归对决策树方法本身进行改进。

#### 参考文献:

- [1] 吴雄军,蒋官澄,赵琳,等.径向神经网络优化及在储层敏感性定量预测中的应用[J].油气地质与采收率,2012,19(1):107-110.  
Wu Xiongjun, Jiang Guancheng, Zhao Lin, et al. Optimization and application of radial basis function neural network for reservoir sensitivity quantitative forecasting[J]. Petroleum Geology and Recovery Efficiency, 2012, 19(1): 107-110.
- [2] 方文超,姜汉桥,李俊键,等.基于不确定性研究的油田开发后期指标预测方法[J].油气地质与采收率,2015,22(5):94-98.  
Fang Wenchao, Jiang Hanqiao, Li Junjian, et al. A prediction method for oilfield development indices during later period based on uncertainty research[J]. Petroleum Geology and Recovery Effi-



- ciency, 2015, 22(5):94-98.
- [3] 袁照威,陈龙,高世臣,等.基于马尔科夫-贝叶斯模拟算法的多地震属性沉积相建模方法——以苏里格气田苏10区块为例[J].油气地质与采收率,2017,24(3):37-43.  
Yuan Zhaowei, Chen Long, Gao Shichen, et al. A method of sedimentary facies modeling through integration of multi-seismic attributes based on Markov-Bayes model: An example from Su10 area in the north of Sulige gas field[J]. Petroleum Geology and Recovery Efficiency, 2017, 24(3):37-43.
- [4] Li X Y, Li H Q. A new method of identification of complex lithologies and reservoirs: task-driven data mining[J]. Journal of Petroleum Science and Engineering, 2013, 109(9):241-249.
- [5] Insua L T, Hamel L, Moran K, et al. Advanced classification of carbonate sediments based on physical properties[J]. Sedimentology, 2015, 62(2):590-606.
- [6] Quinlan J R. Induction of decision tree [J]. Machine Learning, 1986, 1(1):81-106.
- [7] Breiman L, Friedman J H, Olshen R A, et al. CART: Classification and regression trees [J]. Encyclopedia of Ecology, 1983, 1(1):582-588.
- [8] 高世臣,袁照威.地震属性在沉积相预测中的方法研究——序贯随机模式识别[J].地球物理学进展,2016,31(3):1066-1072.  
Gao Shichen, Yuan Zhaowei. Seismic attributes in facies prediction—sequential stochastic pattern recognition methods [J]. Progress in Geophysics, 2016, 31(3):1066-1072.
- [9] Luo D J, Liu A J. Kernel fisher discriminant analysis based on a regularized method for multiclassification and application in lithological identification [J]. Mathematical Problems in Engineering, 2014, 2014(6):1-8.
- [10] Muñoz T P, Hernandez J V, Martinez E H. Wavelet transform analysis for lithological characteristics identification in siliciclastic oil fields[J]. Journal of Applied Geophysics, 2013, 98(3):298-308.
- [11] 居培,翟应虎.基于正态分布理论的模糊综合评判法识别岩性[J].科学技术与工程,2012,12(21):5253-5256.  
Ju Pei, Zhai Yinghu. Lithology identification analysis based on normal distribution theory and fuzzy synthetic evaluation method [J]. Science Technology and Engineering, 2012, 12(21):5253-5256.
- [12] Liu D Q, Li Y L, Song S Y, et al. Simulation and analysis of lithology heterogeneity on CO<sub>2</sub> geological sequestration in deep saline aquifer: a case study of the Ordos Basin [J]. Environmental Earth Sciences, 2016, 75(11):1-13.
- [13] Bi L, Ren B Y, Zhong D H, et al. Real-time construction schedule analysis of long-distance diversion tunnels based on lithological predictions using a Markov process[J]. Journal of Construction Engineering & Management, 2015, 141(2):1.
- [14] Khan Z A, Tewari R C. Principal component analysis of lithologic variables in Early Permian Barakar coal-measures, western Singrauli Gondwana sub-basin of central India [J]. Journal of the Geological Society of India, 2012, 79(4):404-410.
- [15] 尹帅,丁文龙,汤婕,等.主成分分析法在致密碎屑岩储层有效裂缝识别中的应用[J].油气地质与采收率,2016,23(6):1-7.  
Yin Shuai, Ding Wenglong, Tang Jie, et al. Application of principal component analysis method to effective fracture identification of tight clastic rock reservoir [J]. Petroleum Geology and Recovery Efficiency, 2016, 23(6):1-7.
- [16] 杨兆栓,林畅松,尹宏,等.主成分分析在塔中地区奥陶系鹰山组碳酸盐岩岩性识别中的应用[J].天然气地球科学,2015,26(1):54-59.  
Yang Zhaoshuan, Lin Changsong, Yin Hong, et al. Application of principal component analysis in carbonate lithology identification of the Ordovician Yingshan Formation in Tazhong area [J]. Natural Gas Geoscience, 2015, 26(1):54-59.
- [17] 钟仪华,李榕.基于主成分分析的最小二乘支持向量机岩性识别方法[J].测井技术,2009,33(5):425-429.  
Zhong Yihua, Li Rong. Application of principal component analysis and least square support vector machine to lithology identification [J]. Well Logging Technology, 2009, 33(5):425-429.
- [18] 梁永光.复杂岩性储层流体识别方法研究[J].长江大学学报:自然科学版,2015,12(29):36-39.  
Liang Yongguang. Method of fluid identification in complex lithology reservoirs based on principal component [J]. Journal of Yangtze University: Natural Science Edition, 2015, 12(29):36-39.

编辑 邹澍滢