

doi: 10.12097/gbc.2023.04.030

基于 Adaboost 算法的沉积微相自动识别 ——以陇东气田 Q 区山西组为例

黄千玲^{1,2}, 赵军龙^{1,2*}, 白倩^{1,2}, 许鉴源^{1,2}

HUANG Qianling^{1,2}, ZHAO Junlong^{1,2*}, BAI Qian^{1,2}, XU Jianyuan^{1,2}

1. 西安石油大学地球科学与工程学院, 陕西 西安 710000;

2. 西安石油大学陕西省油气成藏地质学重点实验室, 陕西 西安 710000

1. School of Earth Science and Engineering, Xi'an Shiyu University, Xi'an 710000, Shaanxi, China;

2. Shaanxi Key Laboratory of Oil and Gas Reservoir Geology, Xi'an Shiyu University, Xi'an 710000, Shaanxi, China

摘要: 在油气田开发中, 沉积微相识别对于明确沉积背景及单砂体刻画起着重要的作用。陇东气田地质条件复杂, 主力气藏深度大、产层单一, 仅山 1 段底部产气, 对于多种资料交叉共同分析沉积微相, 仅依靠人工判别沉积微相, 过程复杂且容易出错, 很难在沉积微相和测井数据之间建立精确的对应关系。为了充分利用测井资料, 提高沉积微相划分的效率, 提出一种基于 Adaboost 算法的沉积微相自动识别方法, 为后期气田开发沉积背景及单砂体刻画提供更准确的依据。在研究中, 对测井曲线进行优选, 并进行预处理, 运用数学统计法提取了 6 个特征参数作为训练的输入集, 把沉积微相的类型作为训练的输出结果标签, 从已解释的沉积微相数据中选取共 1210 组作为训练样本, 其中组建的训练样本共约 968 组, 组建测试样本 242 组。研究结果显示, 应用该方法的训练效果和测试结果的准确性分别达到 96.45%, 90.4%, 可以验证该方法在陇东气田 Q 区应用效果较好。

关键词: 沉积微相; Adaboost 算法; 测井; 自动识别; 陇东气田

中图分类号: P539; P618.13; TE151 **文献标志码:** A **文章编号:** 1671-2552(2024)04-0658-09

Huang Q L, Zhao J L, Bai Q, Xu J Y. Automatic recognition of sedimentary microfacies based on Adaboost algorithm: Taking the Shanxi Formation in Q zone of Longdong gas field as an example. *Geological Bulletin of China*, 2024, 43(4): 658-666

Abstract: In oil and gas field development, sedimentary micro-phase identification plays an important role in clarifying the sedimentary background and single sand body delineation. Only the bottom of Shan 1 section produces gas. For the analysis of sedimentary microfacies through multiple data intersections, relying solely on manual discrimination of sedimentary microfacies is a complex and error prone process, making it difficult to establish an accurate correspondence between sedimentary microfacies and logging data. Therefore, in order to make full use of the logging data and improve the efficiency of sediment microphase delineation, this paper proposes an automatic identification of sediment microphases based on Adaboost algorithm to provide a more accurate basis for sediment background and single sand body delineation for later gas field development. In the study, optimization and preprocessing of logging curves were carried out, and six feature parameters were extracted using mathematical statistical methods as the input set for training. The type of sedimentary microfacies was used as the output result label for training, and a total of 1210 groups are selected as training samples from the interpreted sedimentary microphase data, of which a total of about 968 groups of training samples are formed and 242 groups of test samples are formed. The results of the study show that the accuracy of the training results and test results of the application of the method reaches 96.45% and 90.4%, respectively, which can be verified that the method is better applied in Q area of

收稿日期: 2023-04-18; 修订日期: 2023-07-17

资助项目: 陕西省自然科学基金基础研究计划《基于生烃期古构造的致密油有利区识别方法——以鄂尔多斯盆地为例》(编号: 2019JM-359)

作者简介: 黄千玲(1998-), 女, 在读硕士生, 从事测井地质综合研究、测井资料处理与解释及矿物学、岩石学、矿床学研究。E-mail: h1326981095@163.com

* 通信作者: 赵军龙(1970-), 男, 博士, 教授, 从事测井资料处理与解释、复杂油气藏测井评价的工作。E-mail: zjl1970@163.com

Longdong gas field.

Key words: sedimentary microphase; Adaboost algorithm; well logging; automatic identification; Longdong gas field

沉积微相是一个亚相区的最小单元,在剖面上具有特有的结构、构造、厚度、韵律性及沉积特征。沉积微相在油气勘探中发挥着重要作用,对于完善剩余富油区的产量预测非常重要。在传统的沉积微相划分中,将古生物学和沉积学理论与主要岩心孔的面状特征相结合,再结合测井曲线的形态特征来分析岩石厚度和颗粒大小,从而确定沉积物类型,划分沉积微相(Li et al., 2018; 苏建龙等, 2020)。然而,交叉分析多个数据源的过程是复杂的且容易出错的,因此很难准确确定沉积物面和测井数据之间的定性映射。

近年来,机器学习在沉积微相的智能识别方面得到了广泛运用。罗仁泽等(2022)针对测井-沉积微相,提出了一种基于特征构造(DMC)和双向长短期记忆网络(BiLSTM)的沉积微相智能识别方法。罗歆(2023)以东营凹陷北带 Y920 区沙四上亚段砂砾岩体为例,把 K 最近邻分类算法(KNN)和卷积神经网络(CNN)的 4 种纹理参数与取心段不同沉积微相 FMI 图像作为学习样本,训练机器学习和深度学习网络开展沉积微相的分类和识别。邱晨(2023)运用基于 K-means 的贝叶斯判别法建立了对微相细分类型的测井精细识别方法。陈靖(2022)从测井资料入手,通过对主成分的分析降维,形成一个综合的识别参数,并将该参数与岩性测井数据相结合,对沉积微相进行定量识别。

然而,这些机器学习算法在不同区域的使用仍具有一定的局限性,在很多情况下无法满足沉积微相识别的要求。BP 神经网络虽然能够实现输入和输出映射,具有较强的非线性映射能力,但是, BP 神经网络收敛缓慢,容易出现局部小型化的问题,导致网络训练失败,而且在学习时无法观察训练过程,结果难以理解(韩文龙等, 2016); 而 SVM 算法在分类上对大量的训练样本模型进行分类时很困难,并且在调整参数时候较麻烦,手续繁琐。模糊聚类分析对噪声和异常值很敏感,但其结果不稳定(赵忠军等, 2016); KNN 算法的计算复杂度太高,当样本出现不平衡时,对少量样本预测精度较低(李艳华等, 2017); 贝叶斯算法对部分分类问题是有效的,在预测测试样本时简单快速,但它是基于分布独立性的

假设,在现实中预测问题时做到完全独立可行性太低(Bhattacharyas et al., 2018)。此外,这些算法高度依赖于具体的地质和沉积环境,局限性大,适用性不高。

为此,本文提出了一种基于 Adaboost 算法的沉积微相识别方法,对研究区的沉积微相进行识别。首先根据少量具有微相标签的测井数据确定对应的微相类别,对关键微相特征集进行挑选,并对特征集开展训练学习,训练弱分类器,选取恰当的阈值,选择最佳弱分类器,最终将每次迭代训练选出的最佳弱分类器构建成为强分类器(曹莹等, 2013),实现各种沉积微相的正确识别,提高了单井识别沉积微相的效率。

1 研究区沉积微相研究难点

目前关于沉积微相的研究方法很多,其中包括传统地质学方法、地质统计学、地球物理方法等(陈欢庆等, 2008)。

(1)应用基础地质学的方法研究沉积微相。如利用岩心、露头等地质资料研究沉积微相。岩心和露头地质资料是对地表地质条件最直观反映。构型分析法的研究也是利用地质学方法研究沉积微相非常活跃的领域,这种方法的主要问题是数据有限、劳动力负担大、使用范围受限。

(2)各种计算机统计学、数学地质统计方法和数学方法研究沉积微相也是一种常见的研究手段。这包括一般的地质统计学方法和计算机技术与数学相结合所开发的不同建模软件。目前,各种地质统计学方法在中国沉积微相研究中发挥着重要的作用。同时,随着计算机技术等相关学科快速发展,不同建模软件也在不断更新,主要在不同的地质统计学思想上体现。

(3)不同地球物理方法和技术在沉积微相研究中的运用。目前使用最多的是测井和地震数据。研究区本溪组缺失,太原组沉积厚度较薄,山西组沉积厚度增大,总体来说,山西组是海陆转换期,物源方向复杂。传统地质学的研究方法需要的资料有限,现有资料只能起到限制辅助沉积微相划分的作用,本文主要结合数学统计法、运用测井资料进行机器学习,对测井资料数据进行预处理,运用数学统计法

选取特征参数(平均值、相对重心、平均斜率、突变幅度、突变方差根等),最后运用机器学习进行训练沉积微相划分,使沉积微相的划分更加快速、高效。

2 研究区机器学习算法选取对比

在研究沉积微相划分的问题上,机器学习智能识别有很大的作用。但不同类型的机器学习在沉积微相识别上有各自的利与弊。表 1 对不同机器学习方法的原理、优点及局限性进行了分析。

研究区三角洲前缘各种沉积微相较难分类识别,由于 Adaboost 算法在分类中的训练集样本更偏向于难以识别的样本,对误分的样本能够继续学习,得到更强的分类器,直到样本被正确分类。因此,本文选取 Adaboost 算法对沉积微相进行识别。

3 研究区沉积微相自动识别方法分析与应用

3.1 Adaboost 算法原理

Adaboost 算法的全称是一种自适应提升算法,是一种二元分类器,从弱分类器的线性组合中构建一个强分类器(梁云等, 2023)。弱分类器的性能不需要很好,只需要比随机假设好,就可以依靠它来构建一个高度精确的强分类器,Adaboost 算法在使用实例训练数据集的过程中多次选择关键的分类器集

合,使用适当的阈值选择最佳的弱分类器,最后将每次迭代中选择的最佳弱分类器建立为强分类器。级联分类器的设计是为了降低非感兴趣图像的输出率,同时尽可能保持感兴趣图像的输出率,因此,随着迭代次数的增加,所有非感兴趣的图像样本都不会通过,而有趣的样本总是尽可能地通过(韩玉娇等, 2022)。强分类器计算公式为:

$$F(x) = \sum_{i=1}^T \alpha_i f_i(x) \tag{1}$$

式中, T 为该公式(1)中弱分类器的数量。弱分类器的输出值有 1 或-1 两种结果,分别对应于正和负 2 种样本。分类决策规则为:

$$sgn(F(x)) \tag{2}$$

式中, sgn 是符号函数。与公式(1)一样,强分类器的输出值也为 1 或者-1,对应于正和负样本。弱分类器和它们的权重是通过训练算法得到的。

在训练过程中,每个弱分类器都要依次训练,并提取它们的权重。训练样本也被加权,最初所有样本的权重都是一样的,被前一个弱分类器错误分类的样本将被赋予更多的权重,所以下一个弱分类器将更加关注这些困难样本。弱分类器的权重基于其准确性,准确性越高,弱分类器的权重越高。给定 1 个训练样本 (x_i, y_i) ,其中 x_i 是特征向量, y_i 是类别标签,其值为 +1 或-1。学习算法的过程见公式

表 1 不同机器学习方法分析
Table 1 Analysis of different machine learning methods

机器学习方法	算法原理	优点	局限性
BP神经网络	通过自身的训练,学习某种规则,在给定输入值时得到最接近期望输出值的结果	是具有很强的非线性映射能力和柔性的网络结构	学习速度慢,容易陷入局部极小值,网络层数、神经元个数的选择没有相应的理论指导,网络推广能力有限
K-means 算法	输入聚类个数 k 及包含 n 个数据对象的数据库,输出满足方差最小标准 k 个聚类的一种算法	根据较少的已知聚类样本的类别对树进行剪枝确定部分样本的分类,具有优化迭代功能,针对部分小样本可以降低总的聚类时间复杂度	K 值的选定难,初始聚类中心的选择不准确,当数据量大的时候,算法消耗时间久
贝叶斯算法	利用已知的先验概率和新的观测数据来更新后验概率,最后实现分类的	处理高维数据和缺失数据,具有良好的泛化能力和鲁棒性	是基于独立性假设,现实中预测问题很难完全独立
支持向量机	对数据进行二元分类的广义线性分类器	在高维空间中非常高效,即使在数据维度比样本数量大的情况下仍然有效	对大量的训练样本进行分类很困难,而且参数调整的过程很繁琐
Adaboost 算法	是一种二元分类器,从弱分类器的线性组合中构建一个强分类器	训练集样本的权重更偏向于难以识别的样本,为了得到下一个最弱的分类器,对被误分类的样本进行进一步的学习,直到样本被正确分类	得到强分类器的过程需要时间

(3)~(7)。

初始化样本权重值, 所有样本的初始权重相等:

$$w_i^0 = 1/l, i = 1, \dots, l \quad (3)$$

循环, 对 $t=1, \dots$, 依次训练每个弱分类器:

训练一个弱分类器 $f_t(x)$, 并计算其对一组训练样本的错误率 e_t ,

计算弱分类器的权重:

$$\alpha_t = \frac{1}{2} \log((1 - e_t)/e_t) \quad (4)$$

更新所有样本权重:

$$w_i^j = w_i^{j-1} \exp(-y_i \alpha_t f_t(x_i)) / Z_t \quad (5)$$

其中, Z_t 为归一化因子, 是所有样本的权重之和:

$$Z_t = \sum_{i=1}^n w_i^{t-1} \exp(-y_i \alpha_t f_t(x_i)) \quad (6)$$

结束循环, 最后得到强分类器:

$$\text{sgn}(F(x)) = \text{sgn}\left(\sum_{t=1}^T \alpha_t f_t(x)\right) \quad (7)$$

从 Adaboost 算法可以看出, 其包括基于训练集的大小用均匀分布初始化采用权重开始, 然后经过算法的迭代, 在后面的操作中用一个公式来改变和标准化样本权重。一个被错误分类的样本会导致权重增加, 正确分类的权重会相应减少, 表明被错误分类的训练集会在一次次过程中拥有更高的权重。因此, 下一轮训练集样本的权重偏向于更难以识别的样本, 为了获得下一个最弱的分类器, 应通过对被误分类的样本进行进一步的训练, 直到样本被正确分类。当训练达到一定的次数或预期错误率时, 一个强的分类器就产生了(张苗燕, 2018)(图 1)。

3.2 研究区沉积微相划分

山西期是陆地-表层海洋背景下的由陆相向海相过渡的沉积阶段, 开始演化出大面积的陆海过渡三角洲沉积。据前人研究, 研究区主要发育三角洲前缘(夏辉等, 2022)。利用测井曲线的形态变化来反映沉积旋回和沉积背景, 与测井曲线的幅度、顶底接触关系、光滑程度、形态、齿中线等有密切关系(表 2)。

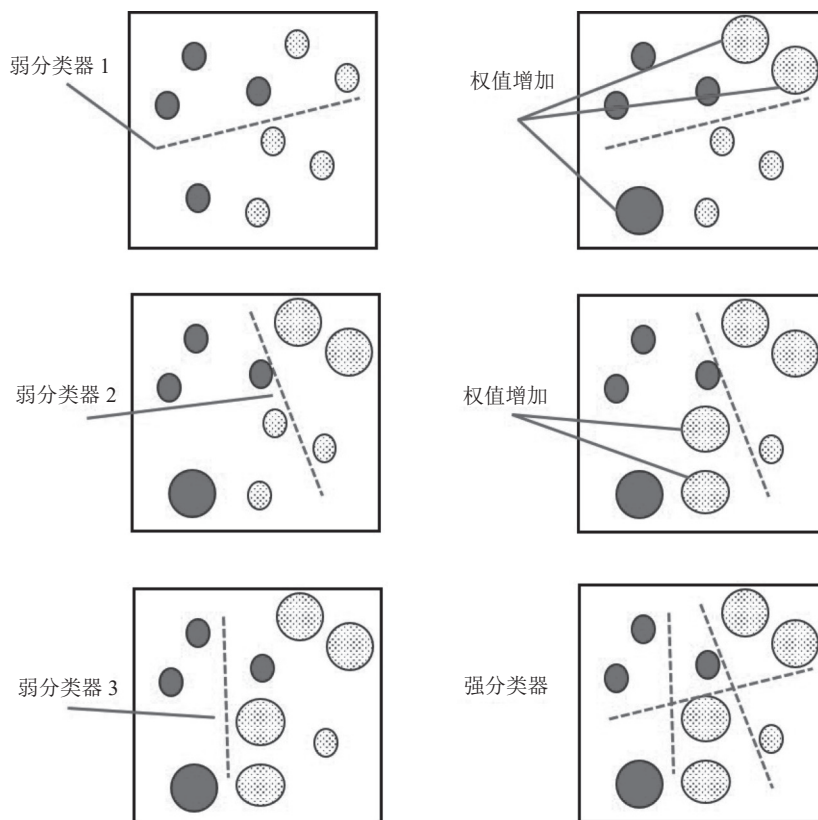


图 1 Adaboost 算法训练过程图

Fig. 1 Adaboost algorithm training process diagram

3.3 测井曲线预处理及特征参数优选

3.3.1 测井曲线的预处理

首先对原始测井曲线数据进行了传统滤波处理,采用滑动平均数字滤波法,选用 9 个窗口,也就是 9 个数据点,对应测井数据中为 $0.125 \times (9-1)=1\text{ m}$ 中的 1 m,对研究区的 20 口井的部分测井曲线进行了滤波处理。图 2 为研究区 Q1 井的 GR 和 SP 曲线滤波处理。可以明显看出,自然伽马(GR)曲线在进行滤波处理后,去除曲线的毛刺,使曲线更加平滑,更加方便对测井曲线数据进行统计。

3.3.2 特征参数的选取

从测井曲线中提取代表沉积微相形态的特征参数是量化沉积物微观形态的关键(王宇喆, 2017)。

在本次研究中,测井曲线的形态参数被用于沉积物微层的定量识别,而选择合适的测井曲线是重点。自然电位与伽马曲线对岩性反映灵敏,可以很好地反映沉积环境,深浅侧向电阻率可以很好地反映目的层物性及电性特征。因此,选择自然电位(SP)、自然伽马(GR)和深、浅侧向电阻率曲线作为本次研究的对象,并统计了各类沉积微相对应测井值范围(表 3)。

测井曲线的数值范围可以用来反映沉积微相的特征参数(国景星等, 2018)。因此,为了更好地比较特征参数,对测井曲线进行归一化处理,使测井曲线数据的值在 0 和 1 之间。

表 2 研究区三角洲前缘亚相的伽马测井曲线类型

Table 2 Gamma logging curve types of delta front subfacies in the study area

曲线形态	箱形	漏斗形	钟形	下部漏斗形,上部为钟形	呈低幅微齿状
微相类型	水下分流河道	河口坝	分流河道侧翼	前缘席状砂	分流间湾
曲线					

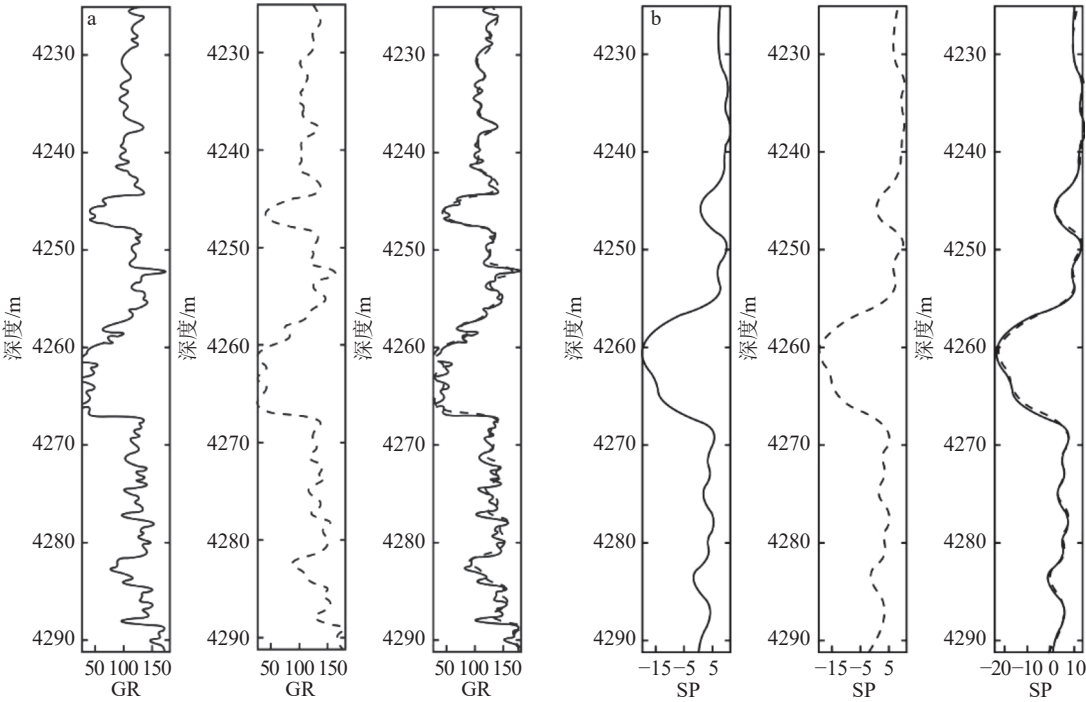


图 2 研究区 Q1 井的 GR 和 SP 曲线滤波处理图

Fig. 2 Filter processing of GR and SP curves of well Q1 in the study area
a—GR 前后滤波对比图; b—SP 前后滤波对比图

表 3 山 1 段三角洲前缘相测井曲线数值范围

Table 3 Numerical range of logging curves for delta front facies in Shan 1 section

沉积微相	测井曲线数值范围			
	自然电位/ mV	自然伽马/ API	深侧向电阻率/ ($\Omega \cdot m$)	浅侧向电阻率/ ($\Omega \cdot m$)
水下分流河道	-30.4 ~ -20.3	24.3 ~ 52.1	15.17 ~ 63.14	14.25 ~ 66.34
河口坝	-20.6 ~ -6.8	68.5 ~ 124.6	14.99 ~ 25.02	15.19 ~ 24.87
席状砂	-25.3 ~ 2.7	50.6 ~ 108.2	8.31 ~ 63.85	8.38 ~ 56.15
分流河道侧翼	-15.7 ~ 6.9	43.2 ~ 118.5	23.77 ~ 220.32	20.96 ~ 224.814
分流间湾	10.4 ~ 30.7	101.3 ~ 168.5	13.67 ~ 31.24	12.93 ~ 28.91

$$Y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

式中, Y 为归一化处理后的数据; x 为原始测井数据; $x_{\max}$ 是曲线的最大值; $x_{\min}$ 是曲线的最小值。最大值、最小值的选取是对研究区的山 1 段井段进行选取, 在自动选取的基础上, 对归一化结果不合理的部分给予人工处理干预。

根据前期研究可知, 沉积微相与测井曲线的幅度、形状、测井曲线顶底的接触关系及曲线的光滑程度相关。下列 5 种特征参数选取可以反映这些特征。

(1) 测井平均值

主要反映曲线的幅度大小。幅度大小可以反映沉积物的分选性、粒度及泥质含量, 由此得出沉积环境的变化。

$$V = \frac{1}{N} \sum_{i=1}^N Y_i \quad (9)$$

式中, V 为测井归一化后平均值; Y_i 为测井曲线第 i 个深度点的归一化值; N 为井段内的深度点的数量。

(2) 相对重心

主要反映曲线的形态轮廓, 形态特征可以反映沉积物岩性和力度的垂向变化, 也可以间接反映物源与当时河道变化的情况。考虑到每个钻井段的不同长度, 采用相对重心, 即绝对重心除以钻井段的深度点数量。相对重心较低的曲线 ($W < 0.5$) 为钟形, 相对重心较高的曲线 ($W > 0.5$) 为漏斗形, 重心集中的曲线 ($W = 0.5$) 为箱形或平直。

$$W = \frac{\sum_{i=1}^N iY(i)}{N \sum_{i=1}^N Y(i)} \quad (10)$$

式中, W 为相对重心。

(3) 平均斜率

主要反映曲线形态变化的快慢, 结果显示了上部和下部的接触关系。顶部渐变表明物源供应逐渐

减少并至中断, 底部渐变表明物源供应持续增加, 反映出当时的物源供应及河道变化速度。可用平均斜率来获得曲线与水平线之间的角度。对于增长测井曲线, 如果角度小于 80° , 为钟形, 如果角度在 $80^\circ \sim 100^\circ$ 之间, 为箱形或平坦, 如果角度超过 100° , 则为漏斗形。

$$K = \frac{\sum_{i=1}^N (Y(i) - V)(D(i) - \bar{D})}{\sum_{i=1}^N (Y(i) - V)^2} \quad (11)$$

$$\bar{D} = \frac{1}{N} \sum_{i=1}^N D(i) \quad (12)$$

式中, K 为平均斜率; D_i 为测井曲线第 i 个深度点的深度; $\bar{D}$ 为平均测井深度。

(4) 突变幅度

主要反映曲线形态变化的趋势, 该趋势主要反映沉积过程中垂向变化的突变情况。一般数值越大, 表明曲线的形态为突变, 反之则为渐变。

$$Z = Z_{\max} - Z_{\min} \quad (13)$$

式中, Z 为突变大小; $Z_{\max}$ 为突变部位的最大幅度; $Z_{\min}$ 为突变的最小幅度。

(5) 变差方差根

主要反映整体的波动程度和曲线的平滑度, 间接反映沉积环境的稳定性。一般测井曲线越光滑, 表明当时的物源充足、河道稳定, 测井曲线齿化表明河道不稳定, 存在变化。GS 值越高表示波动频率越高, 平滑度越低, 而 GS 值低表示曲线较稳定, 平滑度越好。

$$GS = (G_1 + S^2)^{\frac{1}{2}} \quad (14)$$

$$G_1 = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} (Y(i) - Y(i+h))^2 \quad (15)$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y(i)-V)^2 \quad (16)$$

式中, GS 为变差方差根; G_1 是反映测井数据在整个微观阶段的局部变化的方差函数; S^2 是反映测井数据在整个微观阶段的整体变化的方差; $M(h)$ 为步长为 h 的数据对 $[Y(i),Y(i+h)]$ 的数量。

3.4 实际资料处理与分析

3.4.1 样本选取

对于 Q 区山西组山 1 段发育的三角洲前缘亚相,共选取 1210 组已解释的沉积微相数据作为训练样本,其中,水下分流河道微相数据 280 组,河口坝微相数据 200 组,席状砂微相数据 210 组,分流河道侧翼微相数据 260 组,分流间湾微相 260 组;组建的训练样本总共约 968 组,测试样本约 242 组。

3.4.2 特征参数提取

在本次研究中,选择自然电位(SP)、自然伽马(GR)和深、浅层侧向电阻率的记录曲线与沉积微相相关,并首先将曲线归一化。从预处理的测井曲线中提取特征参数,包括测井平均值、相对重心、平均斜率和方差根。提取的参数被用作输入数据集,沉积微相解释的结果被用作输出数据。

根据以上特征参数,计算研究区的自然电位(SP)、自然伽马(GR)和深、浅侧向电阻率测井数据,得到山 1 段 5 个沉积微相测井参数的特征参数值。由于 4 条曲线的分析方法相同,本文只对自然伽马曲线进行阐述(表 4)。

3.4.3 Adaboost 算法的训练与测试结果

(1)训练结果

组建训练样本 968 组,其中水下分流河道微相数据 224 组,河口坝微相数据 160 组,席状砂微相数据 168 组,分流河道侧翼微相数据 208 组,分流间湾微相 208 组。

表 4 山 1 段各类沉积微相自然伽马特征参数平均值

Table 4 Average values of natural gamma characteristic parameters for various sedimentary microfacies in Shan 1 section

沉积微相	测井平均值 V	相对重心 W	变差方差根 GS	平均斜率 K
水下分流河道	0.41	0.48	0.27	0.059
河口坝	0.57	0.63	0.26	0.037
席状砂	0.52	0.47	0.28	0.087
分流河道侧翼	0.51	0.35	0.32	0.027
分流间湾	0.50	0.48	0.26	0.106

以 $n=200$, $learn_rate=1$ 作为参数最终结果,用 Adaboost 算法进行训练的结果显示,使用 Adaboost 算法训练的准确率达到 96.45%,该方法具有良好的训练效果,可用于预测沉积微相。

(2)测试结果

将训练好的 Adaboost 算法模型用于模型测试验证中,组建的测试样本中,水下分流河道微相数据 56 组,河口坝微相数据 40 组,席状砂微相数据 42 组,分流河道侧翼微相数据 52 组,分流间湾微相 52 组,将训练好的训练模型用于测试沉积微相的预测准确性,最终得到测试的准确率,达到 90.4%。图 3 为基于 Adaboost 算法的预测混淆矩阵图,图中标签 0 代表水下分流河道微相,标签 1 代表分流河道侧翼微相,标签 2 代表河口坝微相,标签 3 代表分流间湾微相。从图 3 可知,该算法在河口坝微相和分流间湾微相及分流河道侧翼微相上训练效果有误差。部分测试结果见表 5。

3.4.4 同类算法性能对比

对于组建训练样本 968 组,其中水下分流河道微相数据 224 组,河口坝微相数据 160 组,席状砂微相数据 168 组,分流河道侧翼微相数据 208 组,分流间湾微相 208 组;组建的测试样本中,水下分流河道微相数据 56 组,河口坝微相数据 40 组,席状砂微相数据 42 组,分流河道侧翼微相数据 52 组,分流间湾微相 52 组。分别用支持向量机及 K-means 算法与 Adaboost 算法测试和运算结果进行性能对比。

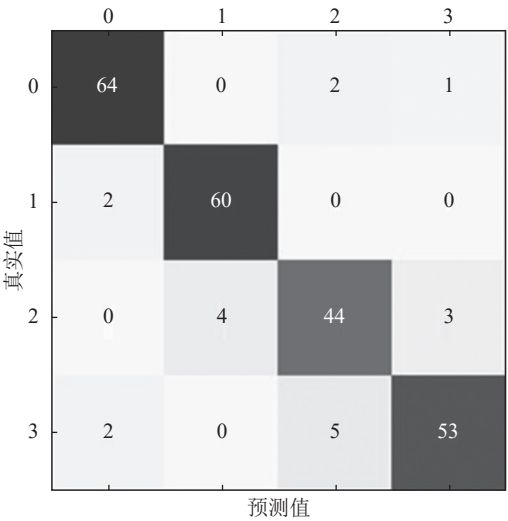


图 3 基于 Adaboost 算法的预测混淆矩阵图

Fig. 3 Prediction confusion matrix based on Adaboost algorithm

表 5 模型部分预测结果
Table 5 Model part prediction results

井名	<i>V</i>	<i>W</i>	<i>GS</i>	<i>K</i>	GR平均值	沉积微相	预测结果
Q1	0.52	0.41	0.30	-0.17	52.06	水下分流河道	水下分流河道
Q10	0.54	0.38	0.29	0.04	70.77	分流河道侧翼	分流河道侧翼
Q10	0.54	0.68	0.32	0.10	110.70	河口坝	河口坝
Q12	0.58	0.63	0.28	-0.03	88.53	河口坝	河口坝
Q12	0.52	0.32	0.91	0.04	81.36	分流河道侧翼	分流河道侧翼
Q12	0.50	0.57	0.24	-0.01	117.54	分流间湾	分流间湾
Q15	0.40	0.56	0.25	-0.01	37.94	水下分流河道	水下分流河道
Q15	0.72	0.60	0.20	-0.11	135.96	河口坝	分流间湾
Q22	0.59	0.66	0.32	0.01	103.27	河口坝	河口坝
Q27	0.47	0.60	0.23	0.08	45.59	水下分流河道	水下分流河道
Q27	0.47	0.60	0.23	0.08	45.59	水下分流河道	水下分流河道
Q27	0.42	0.63	0.23	-0.10	88.64	河口坝	河口坝
Q30	0.29	0.65	0.26	0.03	30.53	水下分流河道	水下分流河道
Q30	0.59	0.42	0.29	0.03	124.30	分流间湾	河口坝
Q20	0.29	0.27	0.28	-0.05	72.14	分流河道侧翼	分流河道侧翼
Q20	0.56	0.53	0.23	-0.08	126.05	分流间湾	分流间湾
Q24	0.49	0.52	0.27	0.05	27.31	水下分流河道	水下分流河道
Q25	0.50	0.59	0.25	0.01	33.47	水下分流河道	水下分流河道
Q26	0.44	0.52	0.27	-0.06	127.18	分流间湾	分流间湾
Q23	0.40	0.64	0.27	-0.04	41.28	水下分流河道	水下分流河道
Q13	0.49	0.60	0.22	0.03	111.67	河口坝	河口坝
Q18	0.42	0.44	0.31	0.01	153.23	分流间湾	分流间湾
Q19	0.18	0.42	0.25	0.02	28.76	水下分流河道	水下分流河道
Q19	0.47	0.59	0.26	0.07	151.07	分流间湾	分流间湾
Q6	0.62	0.61	0.25	0.05	79.87	河口坝	河口坝

(1)支持向量机识别沉积微相

以 $C=20.876$, $\gamma=1.7485$ (C 为系数, γ 为常参数) 为运用支持向量机算法识别沉积微相的判别参数, 运用该算法得到的沉积微相训练结果达到 93.4%, 对测试数据进行测试, 得到测试结果的准确率, 达到 87%。

(2)K-means 算法识别沉积微相

以 $k=4$ (k 为常数) 为运用 K-means 算法识别沉积微相的聚类中心, 使用该算法得到的聚类结果对沉积微相识别的训练结果达到 92.8%, 对测试数据进行测试, 得到测试结果的准确率达到 88.7%。

运用同类算法支持向量机及 K-means 算法对沉积微相进行机器学习识别, 这 2 种算法的训练及测试的准确性都比 Adaboost 算法低。因此, 选取 Adaboost 算法模型在研究区应用更好。

4 结 论

(1)运用 Adaboost 算法对陇东地区三角洲前缘进行沉积微相划分, 明确沉积微相类型, 对测井曲线进行预处理, 基于测井曲线对特征参数选取, 最后选择测井平均值、相对重心、平均斜率及变异方差根等作为特征参数。

(2)选取 Adaboost 算法对沉积微相识别, 得到的训练样本准确率达 96.45%, 测试样本准确率达 90.4%。Adaboost 方法由弱分类器组建的强分类器模型在研究区应用较好, 能够达到应用效果。

(3)对于陇东气田复杂的地质条件, 选取 Adaboost 算法对沉积微相识别, 使过程更加简单不易出错, 沉积微相和测井数据之间的关系更加容易建立, 沉积微相划分更加高效快速。

参考文献

- Bhattacharya S, Mishra S. 2018. Applications of machine learning for facies and fracture prediction using bayesian network theory and random forest: Case studies from the Appalachian basin, USA [J]. *Journal of Petroleum Science and Engineering*, 170: 1005–1017.
- Li L Y, Wei J C, Yin H Y, et al. 2018. Influence of sedimentary facies on reservoir quality and distribution of diagenetic features in the Funing Formation, Wanglong Zhuang oilfield, Subei Basin, Eastern China [J]. *Arabian Journal of Geosciences*, 11: 432.
- 曹莹, 苗启广, 刘家辰, 等. 2013. AdaBoost 算法研究进展与展望[J]. *自动化学报*, 39(6): 745–758.
- 陈欢庆, 朱筱敏. 2008. 精细油藏描述中的沉积微相建模进展[J]. *地质科技情报*, (2): 73–79.
- 陈靖. 2022. 基于录测井资料的沉积微相定量识别方法研究——以渤中 26-X 构造为例[J]. *石油地质与工程*, 36(5): 21–27.
- 国景星, 陈铭. 2018. 基于支持向量机的浊积扇沉积微相自动识别[J]. *甘肃科学学报*, 30(2): 25–31.
- 韩文龙, 王延斌, 高向东, 等. 2016. 基于 BP 神经网络的沉积微相识别[J]. *煤炭技术*, 35(12): 114–116.
- 韩玉娇. 2022. 基于 AdaBoost 机器学习算法的大牛地气田储层流体智能识别[J]. *石油钻探技术*, 50(1): 112–118.
- 李艳华, 王红涛, 王鸣川, 等. 2017. 基于 PCA 和 KNN 的碳酸盐岩沉积相测井自动识别[J]. *测井技术*, 41(1): 57–63.
- 梁云, 门昌骞, 王文剑. 2023. 基于模型决策树的 AdaBoost 算法[J]. *山东大学学报 (理学版)*, 58(1): 67–75.
- 罗歆, 闫建平, 王军, 等. 2023. 基于 FMI 图像深度学习的砂砾岩体沉积微相识别方法——以东营凹陷北带 Y920 区块沙四上亚段为例[J]. *沉积学报*, 41(4): 1138–1152.
- 罗仁泽, 周洋, 康丽侠, 等. 2022. 基于 DMC-BiLSTM 的沉积微相智能识别方法[J]. *石油物探*, 61(2): 253–261, 338.
- 邱晨, 闫建平, 钟光海, 等. 2023. 深层页岩气地层沉积微相细分类型测井识别与应用——以四川盆地泸州地区五峰组-龙马溪组为例[J]. *岩性油气藏*, 34(3): 1–13.
- 苏建龙, 蒲勇, 缪志伟, 等. 2020. 元坝地区灯影组四段丘滩体相控储层预测[J]. *石油物探*, 59(4): 607–615.
- 王宇喆. 2017. Sufyan 油田 AbuGabra 组沉积微相识别方法[J]. *大庆石油地质与开发*, 36(3): 44–48.
- 夏辉, 王龙, 张道锋, 等. 2022. 鄂尔多斯盆地庆阳气田二叠系山西组 1 段层序结构与沉积演化及其控制因素[J]. *石油与天然气地质*, 43(6): 1397–1412, 1488.
- 赵忠军, 刘焯, 王凤琴, 等. 2016. 基于支持向量机的辫状河测井沉积微相识别[J]. *测井技术*, 40(5): 637–642.
- 张苗燕. 2018. 基于 Adaboost 的铁路扣件完损性检测方法研究[D]. 兰州交通大学硕士学位论文.