

DOI: 10.12401/j.nwg.2023029

机器学习算法在地球物理测井中的适用性及应用

程希^{1,2}, 周军³, 傅海成⁴, 罗雄民³

(1. 西安石油大学地球科学与工程学院, 陕西 西安 710065; 2. 西安石油大学院士专家工作站, 陕西 西安 710065; 3. 中国石油集团测井有限公司, 陕西 西安 710075; 4. 中国石油集团勘探开发研究院, 北京 100083)

摘要: 机器学习, 特别是深度神经网络学习算法的发展, 正在改变人们发现知识的方式。目前油气工业正在向转向非常规和深海的油气勘探和开发。基于有限岩石物理参数建立的评价解释模型难以满足反映非常规储层复杂的岩性和结构, 这使传统测井评价技术受到了很大的挑战。以油气大数据为基础、机器学习算法为核心、油气大数据云计算为动力以及油气应用场景为源泉的油气人工智能 (Oil & Gas AI) 极大地改变传统的油气工业各个领域。笔者以地球物理测井为研究对象, 依据数据驱动的地球物理知识发现原理和机器学习属性, 按照“数据-算法-平台-知识-应用场景”研究思路, 开展机器学习算法在测井技术中的适用性研究。对机器算法的内在特性、原理、质量控制、硬件要求, 学习模型选择、测试以及性能评价实现过程进行分析。笔者梳理和总结了机器学习算法在测井中适用性的树状图, 尤其是在油气测井的方法研究、数据处理以及地层评价中的应用潜力与机器学习算法对应关系, 其中包括数据校正的模拟方法、数据标定的岩石物理分析、测井数据质量控制、综合评价以及油藏评价监测。研究表明, 机器学习算法在岩性识别与储层分类、力学评价、以及油藏评价等方面应用有明显的优势, 贯穿于测井方法、仪器设计、测井作业及测井解释中。机器学习算法相对于传统的岩石物理建模方法, 以数据为纽带综合评价岩石物理的多重属性。这从数据科学角度突破了实验条件和物理属性的限制, 具有跨学科、综合性的特点。

关键词: 机器学习算法; 测井大数据; 岩性识别; 油藏评价

中图分类号: P631.81

文献标志码: A

文章编号: 1009-6248(2023)04-0336-13

Applicability and Application of Machine Learning Algorithm in Logging Interpretation

CHENG Xi^{1,2}, ZHOU Jun³, FU Haicheng⁴, LUO Xiongmin³

(1. College of Earth Science and Engineering, Xi'an Shiyou University, Xi'an 710065, Shaanxi, China; 2. Academician Expert Workstation, Xi'an Shiyou University, Xi'an 710065, Shaanxi, China; 3. China National Petroleum Group Logging Co. Ltd, Xi'an 710075, Shaanxi, China; 4. China National Petroleum Group Exploration and Development Research Institute, Beijing 100083, China)

Abstract: Machine learning, especially the development of deep neural network learning algorithms, is changing the way people discover knowledge. As the oil and gas industry is shifting to unconventional oil and gas exploration and development, the evaluation and interpretation model based on limited petrophysical parameters is difficult to meet the complex lithology and structure of unconventional reservoirs, which poses a great challenge

收稿日期: 2022-05-09; 修回日期: 2023-03-03; 责任编辑: 吕鹏瑞

基金项目: 油气藏地质及开发工程国家重点实验室基金项目“基于物理模拟与数据模型的过套管地层真电阻率反演方法研究” (PLN2022JM-14) 资助。

作者简介: 程希 (1971-), 男, 副教授, 主要从事地球物理测井、油气大数据与 AI 科研与教学工作。E-mail: andson@xsyu.edu.cn。

to the traditional logging evaluation technology. Oil & gas artificial Intelligence (Oil & Gas AI), based on oil and gas big data, machine learning algorithms, oil and gas application scenarios, has greatly promoted the application and development of AI technology in various professionals of oil and gas industry. According to the data-driven petrophysical knowledge discovery, and the research idea of the “data–algorithm–platform–knowledge–application scenario”, firstly we analyzed the inherent attributes, principles, quality control, hardware requirements, learning model selection, testing, and performance evaluation implementation process for the machine learning algorithm. The tree graph of the applicability of the machine learning algorithm in logging is summarized, especially the relationship between the application potential and machine learning algorithm in oil and gas logging. These applications include simulation methods for data correction, petrophysical analysis for data calibration, logging data quality control, integrated evaluation, and reservoir monitoring. The study case shows that machine learning algorithms in lithology identification and reservoir evaluation, classification, mechanics, and reservoir evaluation based on the data link across multiple physical properties of petrophysics compared with traditional well logging method, which break through the limitation of experimental conditions and physical properties and has interdisciplinary and comprehensive characterization, had obvious advantages and potentials in well logging technology.

Keywords: machine learning algorithm; logging big data; lithology identification; reservoir evaluation

自从1927年法国斯伦贝谢兄弟获取第一条测井曲线以来,测井技术在经过模拟测井、数字测井、数控测井、成像测井等4代的技术发展,产生了大量的数据。但由于受限于学习算法与硬件算力,这些数据没有很好地被挖掘利用。非常规、超深油气和地热等绿色能源日益引起石油工业界的重视,使传统的测井评价对象也在发生变化。这些非常规油气储层通常表现出低对比度、孔隙结构复杂、矿物成份复杂,以及地热等资源的复杂岩性和评价目的的不同,用传统的岩石物理模型难以满足现场工程需求。近年来,有学者提出了数据模型驱动的地球科学知识发现的思想,开辟了地球科学研究的新思路(Karianne et al., 2019; Markus et al., 2019)。随着CPU、GPU、TPU等硬件算力的显著提升,机器学习、深度学习算法的突破,特别是人工智能算法在图像分类和自然语言处理方面明显的进展,点燃起油气研究人员和现场工程师的热情,他们试图把机器学习应用在测井方法模拟、仪器设计、岩石物理分析、测井作业以及测井解释等方面。一些学者提出基于机器学习与大数据技术的地球物理测井系统的思想,试图形成“测井方法、仪器制造、测井作业、岩石物理以及测井综合解释”一体化,并把人工智能测井(AIL)的发展分为3个阶段(程希等, 2019, 2021)。也有研究人员对包括机器学习在内的人工智能在国内的应用前景进行了综述(邹文波, 2020)。一些学者提出了人工智能技术在测井

技术中的应用构想和人工智能在石油勘探开发领域的应用前景(匡立春等, 2021; 李宁等, 2021)。中国石油天然气集团有限公司正在研发的勘探开发梦想云平台,通过统一数据湖、统一技术平台、通用应用推进油气勘探开发的智能化(杜金虎等, 2020; 赵丽莎等, 2020)。人工智能或机器学习在流体评价中的应用也陆续开展(罗刚等, 2022)。中石化集团公司以油田石化智云工业互联网平台为基础建立智能工厂、智能油田、智能化研究院。中海油集团提出在建设的智能油田技术平台基础上实现智能油田建设、勘探开发数据治理的目标。延长石油则以延长能源互联网平台为基础,实现能源生产、消费、调度,设备状态、数据分析、优化效益、智能仿真、智库服务和绿色经济一体化。地矿系统已经探索结合大数据、云计算、人工智能、5G和区块链等高新信息技术,探索自然资源要素的遥感快速智能识别、地质环境动态变化定量分析与模拟预测、图谱合一的谱遥感地球体检等应用(韩海辉等, 2022; 李志忠等, 2022)。在工程地质方面,有科研人员利用基于高斯混合聚类算法的西安市人工填土空间分布研究(刘梁等, 2022)。在国外,斯伦贝谢公司提出“通过DELFI认知E & P环境发挥机器学习和人工智能的力量”构想,用数字化更深入地对油藏描述,推进油气服务的智能化,他们的技术专家研究了一种使用机器学习来消除传统地震解释方法造成的操作瓶颈,改进断层解释工作流程,应用该流程

成功地提取了地震体上细微的断层位移。贝克休斯公司提出结合录井、油井监测、地层测试,以测井解释等模块为核心构建大数据平台,以促进人工智能技术在石油和天然气行业中应用。国外研究人员和工程师探索把机器学习等机器学习算法应用于复杂岩性识别、孔隙度求取、测井数据质量控制、横波数据提取、随钻电阻率测井模拟等方面开展探索应用(Akkurt et al., 2019; Gupta et al., 2019; Kuvichko et al., 2019; Oruganti et al., 2019; Wu et al., 2019; Xu et al., 2019)。笔者依据数据和物理模型驱动的知识发现范式,按照“数据-算法-平台-知识-应用场景”的思路,基于测井大数据和大数据基础上的机器学习,开展基于测井大数据的机器学习算法特点、属性,以及地球物理测井评价的应用场景分析,以及机器学习与地球物理测井应用场景的对应关系研究,并以实例分析了其在解决复杂岩性识别、储层分类、力学评价及气藏监测等应用。

1 测井大数据与机器学习算法

1.1 测井大数据的形成与特点

近年来,测井大数据和人工智能(AI)在油气工业界受到越来越多的关注,正在成为解决复杂测井评价问题最常用和最有效的方法之一。面向测井大数据的机器学习算法的发展、硬件算力的增强和存储器价格的降低,为机器学习算法的应用提供了数据和硬件支撑。而测井大数据的形成则包括多源化的测井作业,分布式井下传感器,岩石物理分析测试数据,以及与其相关的地质勘探、钻井、录井、测试等数据。测井技术发展过程中形成的多样化记录格式的数据,与测井相关的数据,如录井数据,地质测试,岩心分析,动态测试,模拟数据,后校正和解释处理结果,永久传感器记录数据等形成测井大数据。测井大数据除呈现出大数据的共有特征之外,还具有多尺度性特点,即岩心扫描数据具有微米到厘米尺度,电声成像数据分辨率具有毫米到厘米尺度。常规的微电阻率和声核探测深度在厘米尺度级,而侧向感应类在 2 m 尺度内,过套管电阻率测井则在 10 m 尺度内,井间声波和电磁则在几百米到 1 km 范围内。

1.2 测井机器学习算法的实现与适用性

人工智能是一种机器吸收信息来执行具有人类智能特征的任务和能力的技术,如识别物体形态和声

音、从环境中学习和解决问题。机器学习是对一组数据进行分割、排序和转换,提取问题内在特征,以最大限度地提高对目标数据集中的模式进行分类、预测、聚类或发现的能力。深度学习属于机器学习算法,它构建了越来越复杂的学习层次结构,具有学习的正向和反向传播能量。多层人工神经网络是深度学习算法的例子。

1.2.1 机器学习算法实现

机器学习库是通过云平台的 Spark 接口,并在 Python 和 R 库中与 NumPy 相互操作。可以使用任何分布式大数据源,这样易于在云计算的工作流中实现。以集成机器学习为例,说明机器学习算法的实现。集成机器学习是一种将弱学习算法的结果融合成高质量分类或预测的新方法。集成机器算法的泛化能力通常比单个学习者的泛化能力强。首先,学习数据可能无法提供足够的信息用于单个最佳机器学习算法。其次,单个学习算法的搜索过程很慢。正在搜索的假设空间有可能不包含真正的目标函数,而集成机器学习算法能更好逼近真实函数,尤其是对不稳定的学习算法,如决策树或神经网络。机器学习算法根据对输入数据的人工干预程度,可分为监督学习、无监督学习、半监督学习、强化学习。监督学习主要由决策树、朴素贝叶斯法和支撑向量机等。非监督学习主要是主成分分析,主要用于数据降维以及特性挖掘。

集成学习方法则由提升法和聚类法。假设训练数据为 $\{x_i, y_i\}_{i=1}^N$, 这里 $x_i \in \mathbb{R}^K$, $f_i \in \{-1, 1\}$ 。

当 i 从 1 变化到 N 时,假设给定测井训练数据。假设 f_m 是一些弱分类器。损失函数 I 的值为 0 或 1, 定义为下式:

$$I(f_m(x), y) = \begin{cases} 0, & \text{if } f_m(x_i) = y_i \\ 1, & \text{if } f_m(x_i) \neq y_i \end{cases} \quad (1)$$

找到合适的分类器 m 以降低分类的最小误差:

$$m = \frac{\sum_{i=1}^N w_i^{(m)} I(f_m(x_i) \neq y_i)}{\sum_i w_i^{(m)}} \quad (2)$$

式中: $I(f_m(x_i) \neq y_i) = 1$ 。如果 $f_m(x_i) \neq y_i$, 否则为 0。

$$\alpha_m = \ln \frac{1 - \epsilon_m}{\epsilon_m} \quad (3)$$

其权值系数更新由公式(5)完成:

$$w_i^{(m+1)} = w_i^{(m)} e^{\alpha_m I(f_m(x_i) \neq y_i)} \quad (4)$$

经过学习处理,最终的分类器基于弱分类器的线性组合:

$$g(x) = \text{sign} \left(\sum_{m=1}^M \alpha_m f_m(x) \right) \quad (5)$$

从本质上讲,集成算法是一种贪婪算法,它通过优化一个弱分类器的权值并添加一个弱分类器,逐步地构建一个强分类器 $g(x)$ 。回归算法依据 RMSE 或 MSE 最小化,调整权重来优选学习算法。通过算法试算的 RMSE 比较,其权重函数为:

$$W_n = \begin{cases} 1, & y_i \leq y_{i+1} \\ 0, & \text{其他} \end{cases} \quad (6)$$

例如, y_1 、 y_2 分别为公式(6)中的机器学习算法 1 和 2, 试算出 MAE 或 RME。当第 i 个算法得到的 MAE 偏差小于第 $i+1$ 个算法时,选择第 i 个算法,继续比较,直到选择出,最佳算法为止(图 1)。机器学习算法是通过已有的结果数据构建模型,来预测未知的结果,它是基于已知结果的数据集构建学习模型。通过将这些模型应用于新数据集,并自动完成预测。它擅长从数据中提取特性知识,以构建精确的模型。机器学习算法最适合基于已知结果的观察对新数据进行快速、自动预测。机器学习算法能很方便地测试和优化仿真模型,它通常是通过大数据平台的计算机高级语言来实现的。

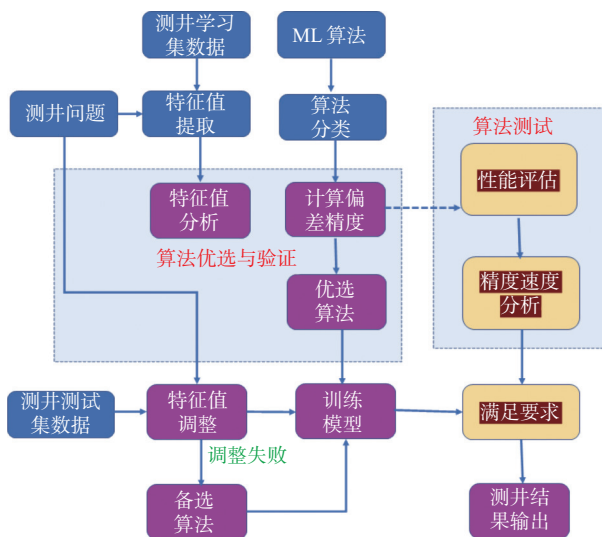


图 1 自适应机器学习算法处理流程图

Fig. 1 Processing chart on adaptive machine learning

现有的测井数据结合其他数据,应用机器学习算法以解决实际测井解释问题。这就需要花费大量时间来选择、运行、调整和算法评估机器学习算法,以确保能有效地利用时间来达到完成测井解释问题的目标。需要评估所用算法的性能和作用,它是训练和测试算法的依据。应用这一指标 RSME(MAE)或精

度对多种算法的结果进行评估,并根据结果可以确定哪些算法的参数需要调整,能取得更高的精度。

机器学习模型的性能是用经过训练的模型对测试数据集预测的度量,例如分类,回归和聚类进行评价。对测井数据进行预处理,训练数据和测试数据要求格式统一。根据分类、回归或者预测问题,选择适合不超过 10 个标准算法,然后按照测试指标评价它们的性能。如果要测试许多方法,则可能必须重新预处理所准备的数据,并减小所选数据集的大小。为了提高效率,尽可能使用较小的数据集对不同的机器学习算法进行评估,而使用数据集中进行算法参数调整,以获取更高的评价指标(图 2)。从总的数据集中,需要选择一部分数据作为训练集,而另一部分数据作为测试集。算法将在训练集数据上进行训练,并在测试集数据上进行评估,可以随机选择数据分割。训练数据集能够体现预测模型的特征。特别是测试数据集与预测数据集数据架构必须一致。与使用测试和训练数据集相比,交叉验证是使用整个转换后的数据集来训练和测试机器学习算法。将数据集分成多个大小相等的实例组,并在所有实例组上训练该模型,在其他测试组数据上测试模型。不断重复这一过程,以使实例数据都有机会作为测试数据集。对性能指标(RSME 或 MAE 等)评估学习算法的性能,也把训练组和测试组数据尽可能的互换,以防止出现过拟合问题。测试各种机器学习算法,就是根据测井问题确定是应用监督机器学习、非监督机器学习。然后,确定测井解释问题的性质以评估优选机器学习算法。

分类对于测井技术而言,主要用于测井岩性识别。集成回归算法的函数为:

$$Y_{\text{new}} = \text{predict}(\text{tree}, X_{\text{new}}) \quad (7)$$

式中: X_{new} 为新数据, predict 为训练学习的函数, tree 为所选的树分类方法。利用交叉验证是为了更好地了解树对新数据的预测准确性。通常,交叉验证随机把训练数据分成 10 个部分它训练了 10 棵新树,每棵树上 9 棵树数据。它检查了树上每个新树的预测准确性训练树中未包含的数据。这种方法给出了很好的估计生成树的预测精度,因为它可以测试新树在新数据上的准确率或精度。集成学习方法的特点是将许多弱学习者的结果融合到一种高质量的整体预测器中,这些算法具有相同的语法,根据需求探索出合适的精度最好的方法。

集成函数(Ens)为:

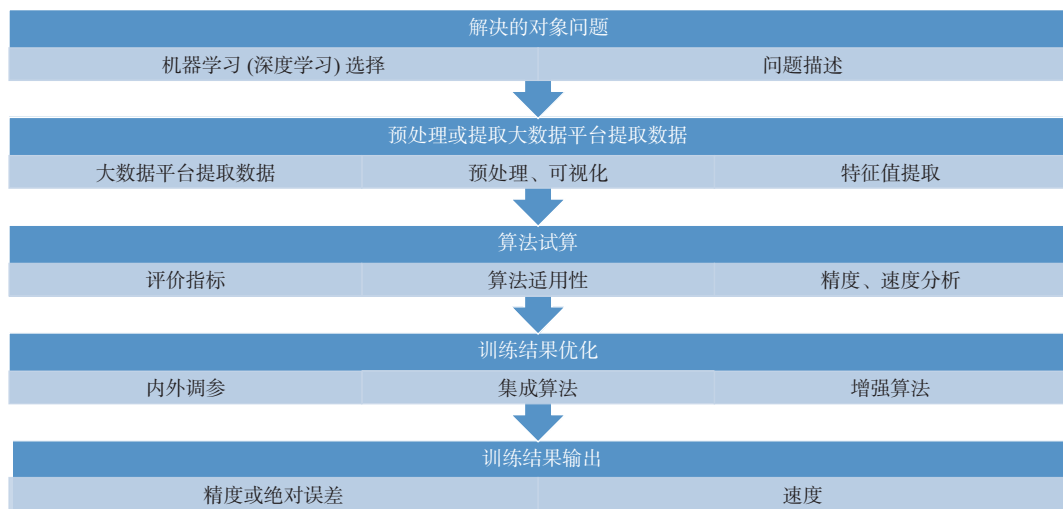


图2 基于机器学习算法的测井解释问题处理流程图

Fig. 2 Logs interpretation flow based on machine learning

$$\text{Ens} = \text{fitensemble}(X, Y, \text{model}, \text{numbers}, \text{learners}) \quad (8)$$

式中: X 数据矩阵, 每一行包含一个观测值, 每一列包含一个预测变量。 Y 为响应, 它具有与 X 中的行相同数量的观察值。 model 是代表使用方法的类型。 numbers 是指每个要素中以 ENS 为单位的弱学习器的数量。因此, Ens 的元素数是学习者中的元素数量。评价学习质量的方法, 首先在独立的测试集上评估整体, 也可以通过交叉验证来评估整体。

1.2.2 适用性

现有的测井技术在储层评价、岩石物理、仪器设计、测井信息采集和模拟仿真等方面, 由于受有限模型和理论模拟的复杂性的影响, 其应用受到了限制。以概率分析、统计学和数据科学为基础的机器学习, 则在解决上述问题上有极大的优势。对测井工业而言, 机器学习算法的应用主要集中在测井仪器的快速物理建模, 油藏油井监测, 测井数据质量控制; 虚拟测井, 数据生成。目前测井技术人员主要关注于困难井的测井数据质量控制、缺少数据修复、多井多条曲线的深度匹配; 多井测井解释中饱和度、渗透率确定、岩性或岩相识别、力学评价、地震反演中声阻抗信息提取; 开发中油井油藏监测、油井寿命和油藏寿命预测。

机器学习算法的特点以及在测井技术上的应用。测井数据质量控制、测井模拟仿真以及岩石物理测试数据, 是测井综合评价分析的基础。综合评价从单井的储层评价, 岩石力学各向异性分析, 到油藏监测, 机器学习算法都能贯穿其中。详细的解决问题、实现手

段、优势以及效果(表1)。通过对回归树、支持向量机、高斯回归、集成树分析等机器学习算法进行了对比, 用集成树的机器学习算法修复了声波横波测井曲线, 并对鄂尔多斯盆地页岩油进行了力学评价。根据已有的测井曲线应用深度神经网络(DNN)等算法产生虚拟测井曲线。以测井曲线为对象、应用神经网络对岩石静态泊松比预测。利用神经网络、梯度回归、高斯回归等方法预测了随钻电阻率测井响应。利用贝叶斯方法对超深方位电阻率测井进行了解释。对声波信号应用主成分分析和支持向量机的方法进行分类。应用机器学习算法对测井数据进行质量控制, 测井曲线深度匹配。根据机器学习算法的特点, 可以应用在地质相分类或储层分类, 岩石物理实验模拟, 地质力学特性预测, 机器学习算法的分类与其在测井中应用的对应关系, 以及机器学习算法在测井中的应用(图3, 表2), 其中, 相同颜色代表树枝分类的相同层(图3)。

集成机器学习算法包括套袋法、提升法、叠加法以及随机森林法。套袋法的目的是降低模型的高方差。以决策树为例, 说明其特点。在每个子样本训练数据的基础上建立多个决策树。在不同的决策树上训练数据, 减少在每个决策树上过拟合的问题。为了提高模型的效率, 每棵决策树都进行深度成长。每个决策树的结果被聚合以形成最终的预测结果, 这样经过聚合, 预测数据的方差就降低了。提升法是按照顺序集成多个的同类型分类器, 每个模型或分类器基于下一个模型将使用的特征值进行测试。它通过平均弱学习模型的权值, 使弱学习模型变成强学习模型。

表 1 机器学习算法在测井中的应用统计表

Tab. 1 Machine learning for logging applications

解决问题	实现手段	优势	效果
复杂岩性识别, 岩性曲线有限	集成学习算法	突破传统岩性曲线多解性限制	识别多种矿物成份致密砂岩, 碳酸盐岩
电测井仪器、核测井仪器正演模拟算法慢	回归算法、深度神经网络	通过已有的模拟数据, 学习出未知的模拟数据	快速实现仪器仿真设计、为测井仪器数字孪生奠定基础
井眼、层厚以及围岩影响校正	回归算法、卷积神经网络	实现同时校正	确定地层的声电核真参数
岩心测试预测, 岩心电镜图片特征提取	卷积神经网络建模、学习	岩石物理参数预测、扫描电镜图像特征提取	孔隙度、渗透率、孔隙类型与矿物结晶识别
测井曲线异常	相关分析	声电核测井曲线特征对比分析	消除仪器或测量引起的误差, 提高测井解释符合率
多矿物复杂岩性识别	分类方法	岩心、录井、钻井、测井数据学习	页岩油气储层, 复杂碳酸盐岩储层
横波数据提取和修复	回归算法	多测井曲线、岩心测试数据学习	页岩油气储层力学参数评价、压裂效果评价
孔隙度、渗透率、饱和度模型	人工神经网络	泥质, 黏土矿物导电性, 束缚水饱和度参数学习	四性参数, 真参数求取
多井归一化, 多参数提取	聚类算法, 降维算法	多井多学科参数学习	多井储层厚度、孔隙度、饱和度提取
页岩油气压裂力学评价、寿命预测、复杂岩性识别	集成机器学习、循环深度神经网络	多学科参数学习、动态监测数据学习	弹性模量、剪切模量、泊松比参数提取, 产能预测

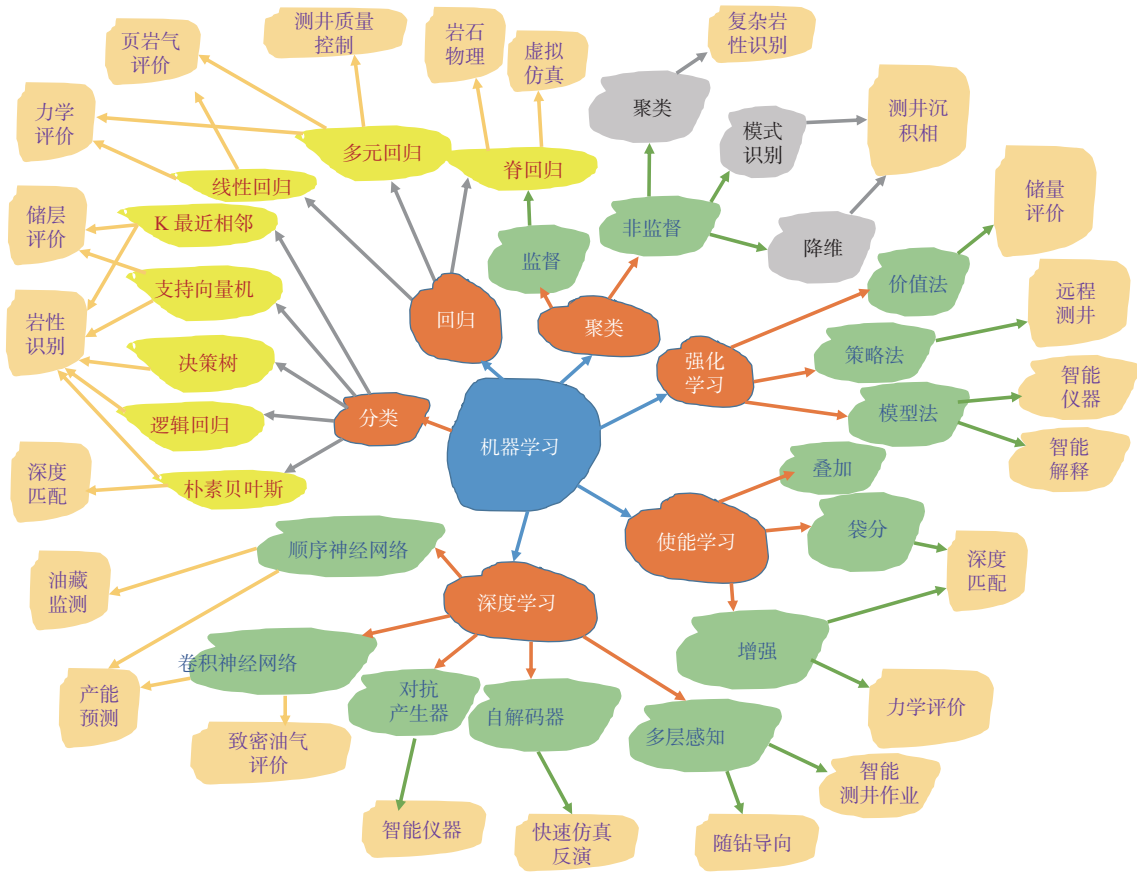


图 3 机器学习算法的分类与其在测井中应用的对应关系图

Fig. 3 Corresponding relationship between the classification of machine learning algorithm and its application in logging

表 2 机器学习算法特点及测井适用性统计表

Tab. 2 Characterization of machine learning and its applicability on well logging

算法	原理	优势与不足	测井适用性
线性回归	在一元的情况下拟合出一条直线，多元时为平面或者曲面，反映数据样本的分类或回归值	速度快、成本低、对于连续线性数据具有可解释性。适合小、中型数据	测井数据修复、测井数据归一化
决策树	决策树是一个由根到叶的递归过程，在每一个中间结点寻找划分属性。若当前结点包含的样本属于同一类别，无需再分支	速度快、结果易于解释、对局部变量敏感。精度有限、误差累积快、不适合于大数据预测	岩性划分、含油性评价、储层质量与分类
随机森林	随机森林就是通过集成的思想将多棵树集成学习的方法，其基本单元为决策树	精度高、适用于高维大数据预测、对数据缺失不敏感。消耗资源大，易出现过拟合	测井沉积相分类、储层分类、含油性评价
朴素贝叶斯法	贝叶斯推理是统计学和机器学习中的一个框架，它使用观察证据来更新假设为真的概率。其关注数据处理的模型的不确定性、编码先验信念和估计错误传播	原理简单，速度快，不受数据噪声和数据缺失影响。有限精度，具有独立的预测器，不适用于大数据。	岩性、岩相划分、含油性评价、储层质量分类、测井沉积相
邻近法	一个样本的属性与数据集中的k个样本的属性最相似，如果这k个样本中的属于某一个类别，则该样本也属于这个类别	非线性。用于中、大型数据	常规测井横波提取、含油性评价、储层质量分类
集成学习算法	包含引导聚合和提升算法。引导聚合是通过将样本训练集合进行随机化抽样，反复抽样训练新的模型。提升算法通过不断地使用一个弱学习器弥补前一个弱学习器不足的过程，实现目标函数值误差小	原理简单、精度高、调节参数少。消耗资源大，适合中小数据量，可解释性差	沉积相与储层分类、岩性划分、数据标准化与修复、深度匹配
主成分分析	利用正交变换把由线性相关变量表示的训练数据转换为少数几个由线性无关变量表示的数据，线性无关的变量称为主成分	中大型数据，针对专业人员	油藏管理监测、历史数据分析、测井知识发现、岩石力学参数、特征值提取
卷积神经网络	卷积滤波器本质上是使用滑动窗口方法的加权向量/矩阵/立方体。根据内核结构，卷积增强了数据的某些特征，例边缘、趋势或平坦区域。卷积是嵌入在神经元水平的卷积神经网络中的，它从之前的层中提取有用的特征	适合于大中型数据。硬件资源要求高	裂缝识别、压裂评价、沉积环境分析
循环神经网络	递归神经网络的信息可以在不同节点之间循环，产生诸如存储器之类的复杂动态	适合于大中型数据。硬件资源要求高	油藏监测、微地震监测、压裂效果评价
长短记忆法	LSTM是使用当前输入和上一个状态传递下来信息，训练得到四个状态。由状态向量乘以权重矩阵之后，再通过激活函数转换成0到1之间的数值，作为一种门控状态。而输入数据通过激活函数将转换成-1到1之间的值	适合于大中型数据。硬件资源要求高	油藏监测、微地震监测、压裂效果评价
深度强化学习	人工智能是一种算法吸收信息来执行具有人类智能特征的任务的能力，例如识别物体和声音、将语言上下文化、从环境中学习和解决问题	适合于大中型数据。硬件资源要求高、数据资源丰富	自主测井、虚拟测井
对抗神经网络	这是一系列无监督机器学习方法，用未知概率密度函数生成样本。生成对抗网络是能区分真实样本和虚假样本的鉴别器网络	计算限制、主观特征选择	学习测井专家知识、智能仪器、数据质量

叠加法结合多种分类或回归技术,使用完整的训练数据集训练较低级模型,然后使用较低级模型的结果训练组合模型。与增强训练不同,每个较低层次的模型都要进行并行训练。来自较低层次模型的预测用作下一个模型的输入,作为训练数据集,并形成堆栈,其中模型的顶层比模型的底层训练得更多。顶层模型具有较好的预测精度,它们是在低层模型的基础上建立起来的。随着堆栈的增加,最后以最小的误差的模型进行预测。组合模型或元模型的预测是基于对不同弱模型或低层模型的预测。它专注于产生更少的偏差模型。随机森林法在自举样本上使用深度树。将每个树的输出进行组合以减少方差。在生长每棵树时,数据集的采样是基于减少不同输出相关性的特征进行树生长的。它适合于具有确定缺失数据的训练集数据。它的每棵树都有不同的结构,这可能导致森林法的偏差略有增加,但由于结果是平均所有来自不同树且特征相关性较少的预测结果,其方差的减少,有利于提高整体的预测性能。集成法是通过分类器的不同组合对复杂数据进行处理,以获得更好的预测或分类效果。在集成学习中,每个模型的预测必定是不相关的。这将使模型的偏差和方差尽可能地低。该模型能在最小误差的情况下提高效率,预测输出。集成法是一种监督学习算法,因为模型之前是用数据集进行训练来进行预测的。在集成学习中,分量分类器的数量应与类标签的数量相同,以达到较高的准确率。

2 测井知识发现

2.1 基于机器学习的测井知识发现

目前的地球物理科学既是以数据驱动,也是以模型驱动的研究领域。地球物理测井参数的反演问题常常将两者联系起来。地球物理测井科学知识的发现将越来越多地来自于对测井大数据的分析、反演理论的新发展以及计算密集型模拟结果。随着地球物理测井过程的计算机模拟在规模和复杂程度上迅速增加,地球物理科学家面临的首要挑战是如何尽可能多地提取有用的信息,如何从数据和模拟以及两者之间的相互作用中获得对新知识的理解。近年来,地球物理科学家应用机器学习这一新技术,从最初集中在大数据分析,已经扩展到使用机器学习(ML)来通过数据驱动的发现和模型驱动的理解,实现对声电核以

及流体之间的耦合过程的更深理解。机器学习具有在更大的函数空间,提取知识特征工程,从而建立复杂的模型,提供了从数据中得出新认识,也是被认为是通向油气人工智能的手段。它涉及到机器可以学习经验,识别数据中的复杂模式和关系。由于ML方法是稳健的、快速的,并能探索一个大的函数空间,为地球物理测井学者提供了能用于在科学数据集中发现新的模式、结构和关系的能力,这些都是通过传统技术不容易发现的。ML可以揭示出以前未被识别的信号或物理过程,并提取关键特征来代表、解释或可视化地球物理测井数据。正演建模和反演时,使用ML进行自动化、建模或反演可能会产生新知识的理解和发现。目前油气工业研究人员,正在探索将机器学习应用在测井模拟、测井作业质量控制、测井解释和储层工程参数评价中。

2.2 基于测井大数据技术与机器学习的知识发现框架

基于测井大数据云计算是实现地球物理测井知识发现的新途径。测井大数据云计算实现了数据存储、文件管理、进程控制、并行任务处理能力。它的分布式系统是一种可伸缩分布式文件系统,可在多台计算机存储数据,具有高度的容错能力。它能实现资源管理功能,并且数据处理分析在集群机上运行,具有并行处理框架,即在主节点接受输入并将其划分为较小的子问题,并将其分配给工作程序节点。它的弹性分布式数据集是一种特殊基础数据结构,能把存储在不同机器上的数据,形成逻辑数据集。测井大数据云计算是提供测井大数据的分布式存储与处理、人工智能学习模型的数据训练、验证和测试的硬件服务框架。其内部包括分布式文件系统,资源管理与调度模块,分布式计算框架,各种数据操作与学习模块,服务连接总线层,以及用户的应用层(图4)。

需要注意的是,随着大数据的用户增加和功能的强大,云计算的数据安全日益引起石油公司的关注。云计算的数据安全采用3种技术手段,保证数据安全。首先选择在云计算平台上使用强密码,也就是必须创建并使用一个强大的密码,以保护数据不被黑客入侵。这些强密码包括:密码长度至少为10个字符。密码必须包含大写字母(A-Z)、小写字母(a-z)、数字(0-9)与特殊字符,例如@、#、\$、%、^、(、)、&、*!。c# a25^ub@2是一个强密码和标准密码。对于不同的帐户使用强而独特的密码,并定期更改密码。其次是

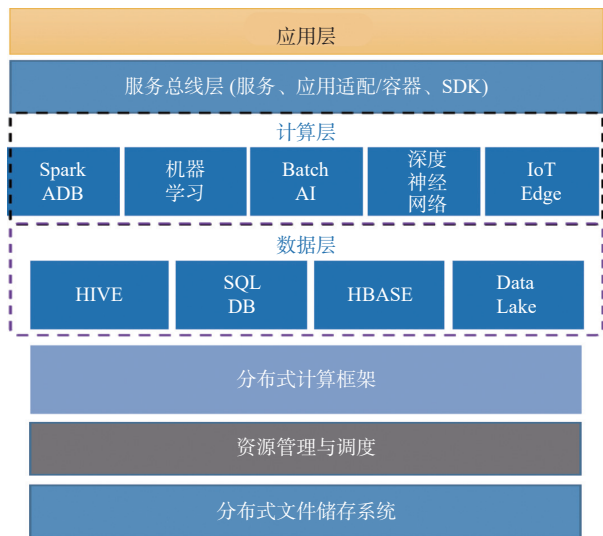


图4 测井大数据云计算模块结构图

Fig. 4 Internal module distribution on private cloud computing

对数据加密,它是云计算中确保数据安全的有效手段之一。它需要用户在访问数据之前对数据解密,这个解密过程也将保护数据免受服务提供商和未经授权用户的网络攻击,其加密和解密的方式包括: AES、MD5、SHA1 等。最后是使用防毒软件,它能帮助保护云数据免于恶意软件,病毒,和不必要的威胁,并防止黑客未经授权访问云数据。

3 应用范例

测井知识发现的实现是通过大数据云计算驱动的机器学习来实现的,以下从岩性识别与储层分类、横波提取及各向异性评价以及油藏评价等方面说明其应用。

3.1 岩性识别

鄂尔多斯盆地的页岩油具有勘探和开发的巨大潜力,但其复杂的储层组成却变得越来越复杂。目前,测井的便捷解释面临更加难以识别岩性,评估和监测储层以及预测产量的问题。仅利用特征的常规的岩性测井曲线来识别页岩未取得良好的效果,因为有些测井响应信噪比低。传统的测井处理和解释工作主要依靠专家的经验,并且效率低。机器学习算法是通过首先将训练数据依靠特征工程,减少为几个可以解释的类,然后每个类学习模型来减少机器学习算法的局限性。该方法可作为连续学习,提取特征值,从而形成自动处理流程来完成测井数据解释。

机器学习算法技术可以使用更多特征曲线来识

别岩性,并且可以使用修复测井丢失的数据。除了岩性测井曲线外,在许多特征曲线,如 AC, DEN, Ra 和 CNL 以及录井、测试等数据用来训练机器学习模型。采用决策树,支持向量机, KNN 和集成算法进行优化学习。结果表明,采用集成法的 Bagging 的训练精度几乎达到 98.5%, 预测值达到 94.3%。图 5 展示了岩性识别声波井径特征的可视化交会图,图 6 和表 3 展示了机器学习算法岩性识别结果以及准确率对比。对所有模型进行了参数选择分析,并将最佳参数用于相应的分类器。集成学习方法变得更具影响力,因为它可以通过使用多种学习算法来补偿弱学习算法的不足以获得更好的性能。在复杂的地质情况下,测井响应与地层参数存在更复杂的非线性函数关系,应用机器学习算法来准确识别岩性,最佳分类器是集成学习,其次是 k 最近邻法,最后是决策树。在用多条曲线进行储层质量分类时,第一组采用 GR, AC, CNL, DEN, RD, RS, RD/RS, RD-RS 学习;第二组采用 GR, AC, CNL, DEN, RD, RS 学习;第三组采用 GR, AC, CNL, RD, RS, RD/RS, RD-RS 学习;第四组采用 GR, AC, CNL, RD, RS;第五组采用 GR, AC 学习。

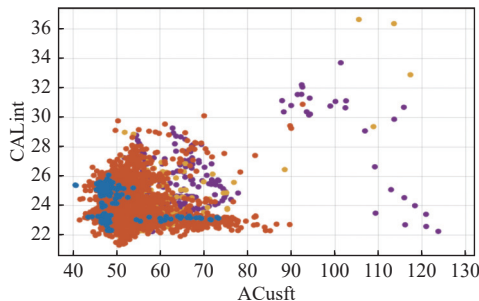


图5 岩性识别声波井径数据的特征可视化图

Fig. 5 Lithology recognition and acoustic well diameter data visual intersection by Machine learning

3.2 横波提取及各向异性评价

当井眼不规则时,井眼内部存在复杂的泥浆流变和天然气涌入时,很难获得可靠的水平横波时差,特别是在软地层中。能应用机器学习算法的方法从各种常规测井曲线中预测横波速度,但该方法除使用纵速度测井作为特征值外,还使用了其他几种测井测量值,如自然伽马,密度,中子,电阻率,孔隙率和饱和度等测井曲线。它是对已有的具有横波速度测井数据集训练,并验证机器学习算法模型。根据测井知识和横波的相关性,在特征选择过程以突出显示哪些测井曲线是横波速度(VS)的良好预测指标。然后训练各

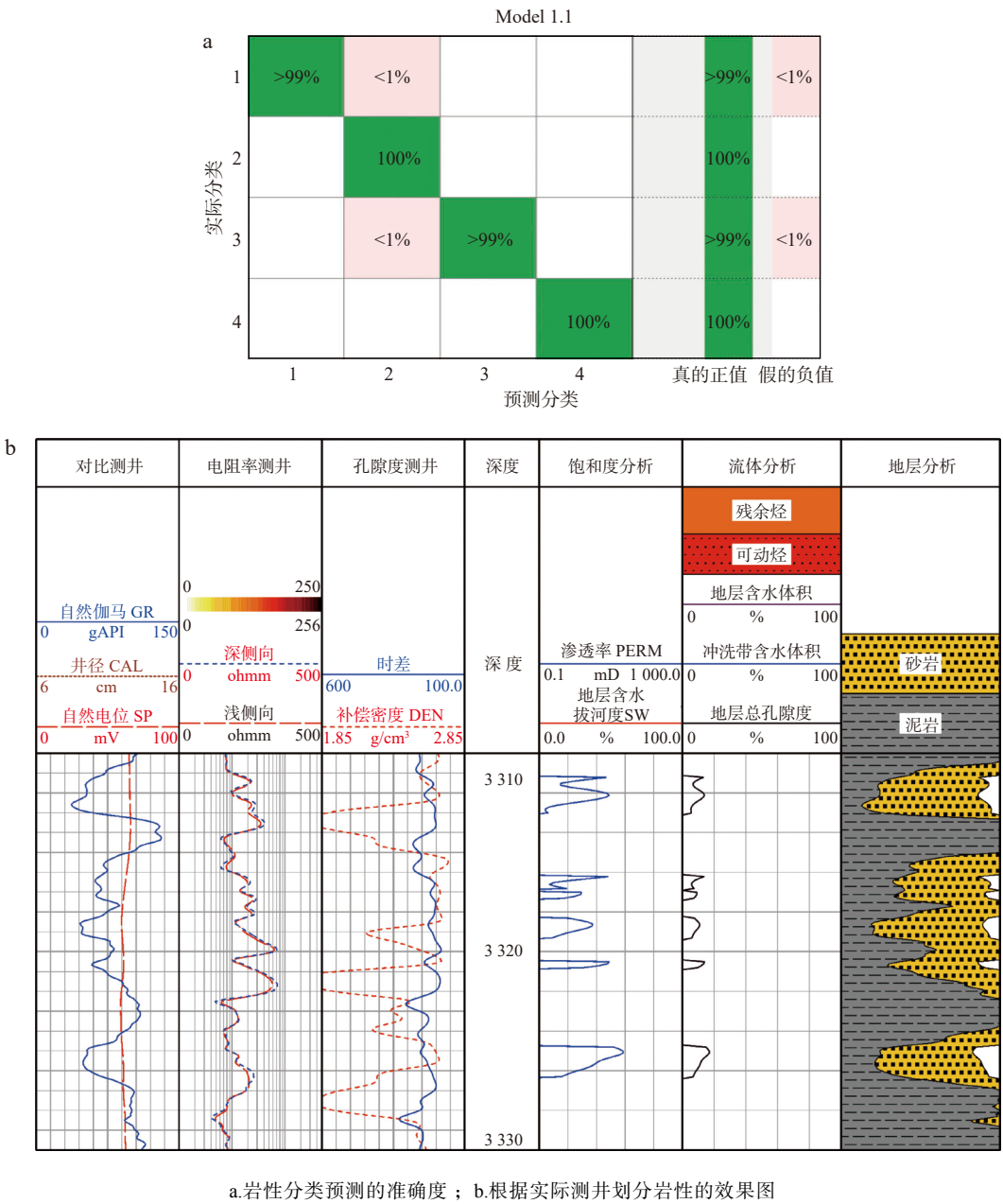


图 6 机器学习算法岩性识别准确率对比图

Fig. 6 The comparison of machine learning lithology identification accuracy

表 3 多种测井曲线下的机器学习算法储层分类结果对比表					
Tab. 3 The comparison of machine learning reservoir classification results under multiple logging curves					
组 (ML算法)	决策树	K邻近法	使能法	深度神经网络	特征曲线数
1	63.8	69.7	89.4	94.6	Pe、Cal、GR、SP、ML、DEN、DT、PHIN
2	66.1	78.6	85.5	93.2	Pe、Cal、GR、SP、ML、DEN
3	61.6	66.7	78.9	91.8	Pe、Cal、GR、SP、ML
4	58.6	64.6	76.7	90.9	Pe、Cal、GR
5	55.9	60.5	72.1	85.4	Pe、Cal

种回归模型,并通过其均方根误差(RMSE)将预测值与各种模型的实际值进行比较,然后选择具有最小RMSE的模型。再对数据集集中的另一口井进行预测,该井用作验证集。

利用机器学习算法中的回归预测修复扩径段横波数据,用于评价致密含气砂岩储层各向异性(图7),对地质力学,岩石物理学和其他应用尤为重要。储层各向异性指数表示储层不同方向矿物颗粒、分层、裂缝、应力等引起横波速度随方向的变化,是用快慢横波速度之差来度量的,可以用以下式来定义:

$$Aniso = \frac{\Delta S}{(S_1 + S_2)/2} \quad (9)$$

式中: S_1 慢横波时差, S_2 快横波时差, $\Delta S = S_1 - S_2$ 。泊松比和杨氏模量用于评价非常规油气储层脆性所必须的2个参数,它可以从偶极声波测井和体积密度测井中估算出来。

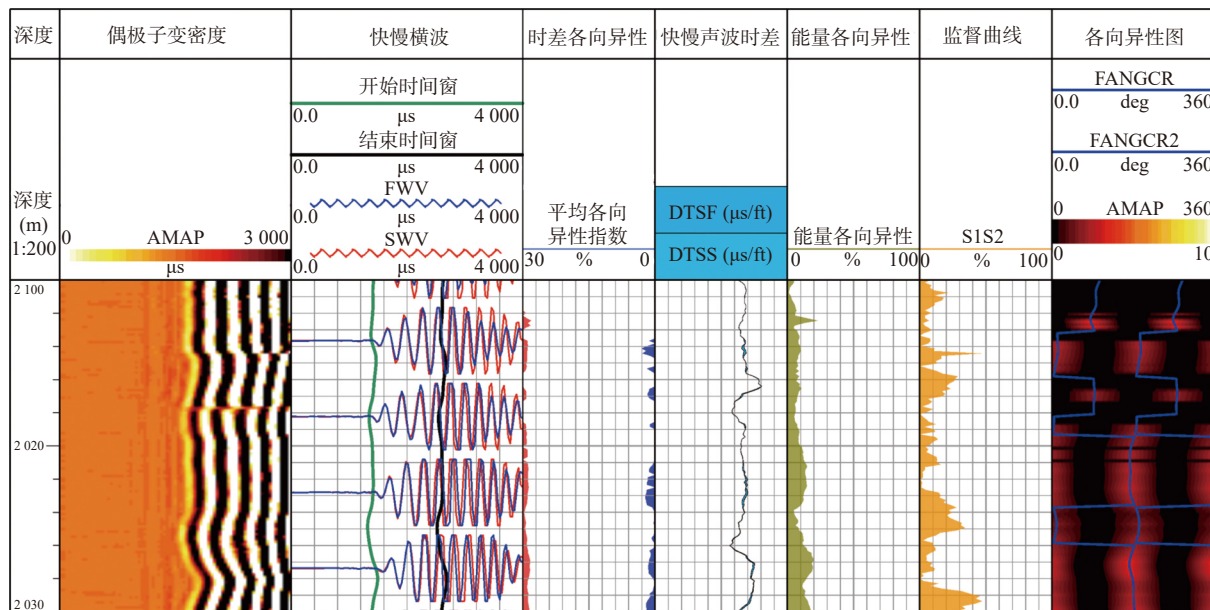


图7 机器学习算法修复横波数据评价致密含气砂岩储层各向异性图

Fig. 7 Evaluation of anisotropy of tight gas-bearing sandstone reservoir using machine learning to recover shear wave data

把已有数据段的数据,分为训练数据和测试数据。利用机器学习算法,建立以纵波时差(速度)、密度、井径以及自然电位等曲线为特征的学习模型。通过优选机器学习算法,得到关于地层横波时差的学习模型。在修复段,利用学习的模型,学习出该段的地层横波参数。图7中第6道的蓝色重叠图代表地层的时差各向异性。

3.3 油气藏评价

从单井扩展到气藏,可以有效了解这个气藏的含气饱和度变化规律,从而降低开发风险。机器学习能解决了更新3D静态模型和油藏参数预测问题。将裸眼井与套管井测井数据进行标准化,深度匹配,应用数字岩心技术,根据测井岩性曲线机器学习相关分析,建立随沉积相变化的可变m、n的阿尔奇含气饱和度计算方法。在套管井测量不同时间段的数据,通过过套管电阻率测井正演模拟,建立地层视电阻率与水泥环厚度、水泥环电阻率、地层参数(厚度、电阻率)以及围岩参数(厚度、电阻率)的响应值。利用实际测井资料,通过机器学习建立地层真电阻率与地层视电阻率与水泥环厚度、水泥环电阻率、地层参数(厚度、电

阻率)以及围岩参数(厚度、电阻率)的学习模型。最后反演出地层真电阻率,实现从地层真电阻率的变化,反映气层以及气藏的含气饱和度的变化,实现气藏开采动态监测,为实现时间序列上的产能变化预测奠定技术基础。

机器学习在气藏监测运行方面具有成本更低,更准确且速度更快的优势。数据驱动的机器学习预测分析已应用于具有大量数据的气藏监测与管理,这些技术包括K最近邻(KNN)、支持向量机回归、核岭回归、随机森林、自适应增强(Adaboost)等。使用机器学习方法进行4D储层物性变化预测的可行性。结合声电模拟算法和过套管电声测井资料,可实现油藏动态监测及油井产能及寿命预测。油藏监测纵向含气饱和度变化分布显示(图8),黄色部分为含气厚度和饱和度在多井纵向上的分布。经过生产测试,在第一、三气藏的井L48第一、二层段发现高产气层,合采的日产气量为3284万 m^3 、水量为3.3 m^3 ;在第一、二、三气藏的井L172井的3个层段有气水同产层,日产气量为3.1万 m^3 、水量为3.4 m^3 ;其他井的含气层段均有气产出。

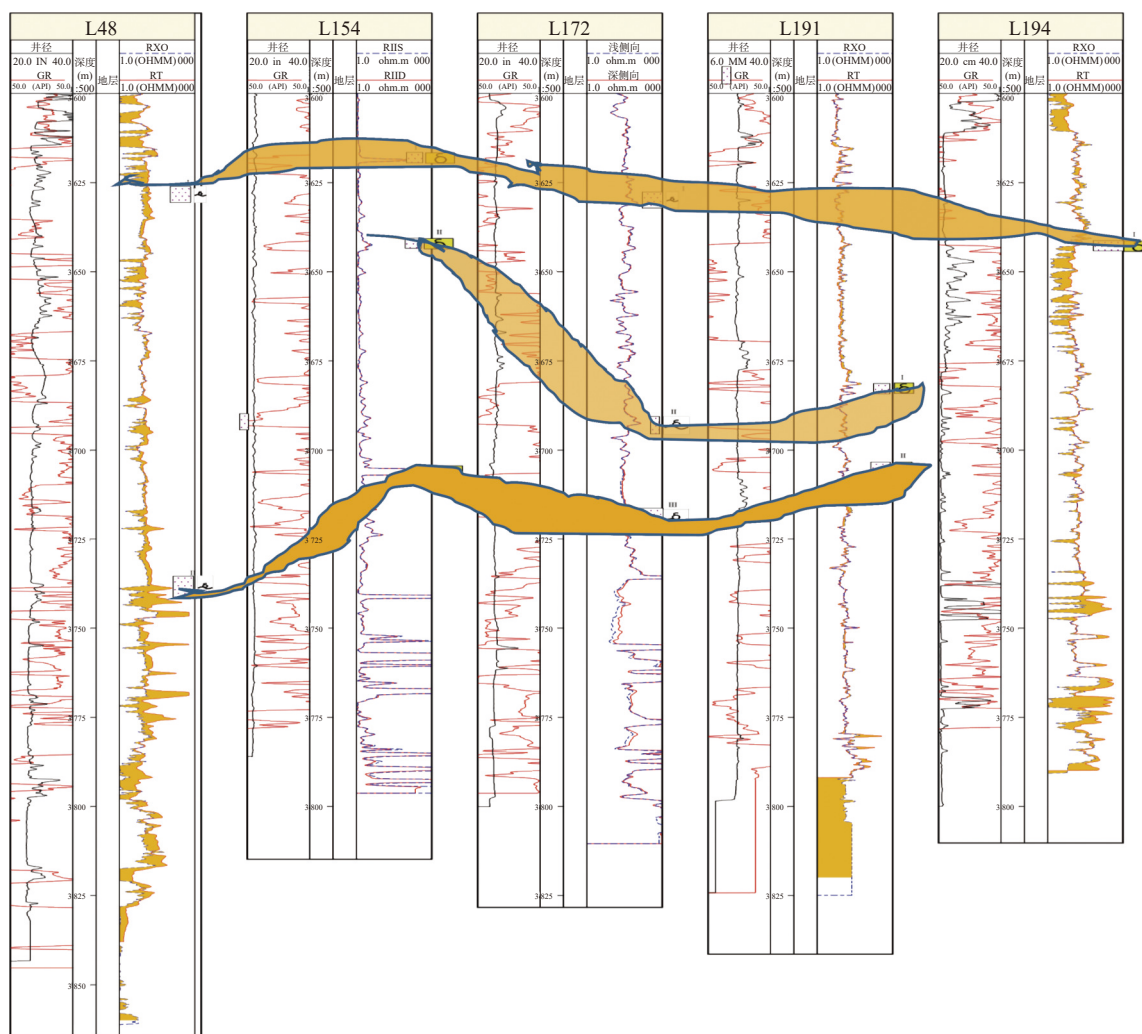


图8 过套管电阻率多井气藏监测图

Fig. 8 Monitoring of gas distribution in multiple wells through casing resistivity

4 结论

(1)按照“数据-算法-平台-知识-应用场景”的研究方法,分析了测井大数据的形成基础;根据机器学习算法的特点和测井技术的分类属性,分析了测井机器学习的实现及在测井中的适用性。

(2)根据基于测井大数据的机器学习知识发现的特点,以及实现框架分析,它在岩性识别、力学各向异性评价以及油气藏监测等方面具有明显的潜力和应用前景。

参考文献(References):

程希,程宇雪,程佳豪,等.基于机器学习与大数据技术的地球物理测井系统[J].西安石油大学学报(自然科学版),2019,

34(6): 108-116.

CHENG Xi, CHENG Yuxue, CHENG Jiahao, et al. Geophysical logging system based on machine learning and big data technology[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2019, 34(6): 108-116.

程希,宋新爱,李国军,等.数据模型与物理模拟驱动的人工智能测井[J].测井技术,2021,45(03): 233-329.

CHENG Xi, SONG Xin'ai, LI Guojun, et al. Artificial intelligence logging driven by data modeling and physical simulation[J]. Logging Technology, 2021, 45(03): 233-329.

杜金虎,时付更,杨剑锋,等.中国石油上游业务信息化建设总体蓝图[J].中国石油勘探,2020,25(5): 1-8.

DU Jinhu, SHI Fuqing, YANG Jianfeng, et al. A general blueprint for upstream business information construction in China[J]. China Petroleum Exploration, 2020, 25(5): 1-8.

韩海辉,李健强,易欢,等.遥感技术在西北地质调查中的应用及展望[J].西北地质,2022,55(3): 155-169.

HAN Haihui, LI Jianqiang, YI Huan, et al. Application and Prospect

- of Remote Sensing Technology in Geological Survey of Northwest China[J]. *Northwestern Geology*, 2022, 55(3): 155–169.
- 刘梁, 石卫, 张晓平, 等. 基于高斯混合聚类算法的西安市人工填土空间分布研究[J]. *西北地质*, 2022, 55(2): 298–304.
- LIU Liang, SHI Wei, ZHANG Xiaoping, et al. Research on Spatial Distribution of Artificial Fill in Xi'an Based on Gaussian Mixture Clustering Algorithm[J]. *Northwestern Geology*, 2022, 55(2): 298–304.
- 李宁, 徐彬森, 武宏亮, 等. 人工智能在测井地层评价中的应用现状及前景[J]. *石油学报*, 2021, 42(4): 508–522.
- LI Ning, XU Binsen, WU Hongliang, et al. Application status and prospects of artificial intelligence in well logging and formation evaluation[J]. *Acta Petrolei Sinica*, 2021, 42(4): 508–522.
- 李志忠, 卫征, 陈霄燕, 等. 新型对地观测技术与地球健康体检[J]. *西北地质*, 2022, 55(2): 56–70.
- LI Zhizhong, WEI Zheng, CHEN Xiaoyan, et al. New Earth Observation Technology and Earth Health Examination[J]. *Northwestern Geology*, 2022, 55(2): 56–70.
- 罗刚, 肖立志, 史燕青, 等. 基于机器学习的致密储层流体识别方法研究[J]. *石油科学通报*, 2022, 7(01): 24–33.
- LUO Gang, XIAO Lizhi, SHI Yanqing, et al. Machine learning for reservoir fluid identification with logs[J]. *Petroleum Science Bulletin*, 2022, 7(01): 24–33.
- 匡立春, 刘合, 任义丽, 等. 人工智能在石油勘探开发领域的应用现状与发展趋势[J]. *石油勘探与开发*, 2021, 48(1): 1–11.
- KUANG Lichun, LIU He, REN Yili, et al. Application and development trend of artificial intelligence in petroleum exploration and development[J]. *Petroleum Exploration and Development*, 2021, 48(1): 1–11.
- 赵丽莎, 史永彬, 金玮, 等. 基于梦想云的测井智能化解释应用研究[J]. *中国石油勘探*, 2020, 25(5): 97–103.
- ZHAO Lisa, SHI Yongbin, JIN Wei, et al. Research on the application of intelligent interpretation of logging based on dream cloud[J]. *China Petroleum Exploration*, 2020, 25(5): 97–103.
- 邹文波. 人工智能研究现状及其在测井领域的应用[J]. *测井技术*, 2020, 44(04): 323–328.
- ZOU Wenbo. Current status of artificial intelligence research and its application in the field of well logging[J]. *Logging Technology*, 2020, 44(04): 323–328.
- Akkurt R, Miller M, Hodenfield B, et al. Machine Learning for Well Log Normalization[C]. Society of Petroleum Engineers, 2019.
- Gupta I, Devegowda D, Jayaram V, et al. Machine Learning Regressors and their Metrics to Predict Synthetic Sonic and Brittle Zones [C]. Unconventional Resources Technology Conference, 2019.
- Karianne J, Paul A, Maarten V, et al. Machine learning for data-driven discovery in solid Earth geoscience[J]. *Science*, 2019, 363(6433): 1299.
- Kuvichko A, Spesivtsev P, Zyuzin V, et al. Field-Scale Automatic Facies Classification Using Machine Learning Algorithms[C]. Society of Petroleum Engineers, 2019.
- Markus R, Gustau C, Bjorn S, et al. Deep learning and process understanding for data-driven Earth system science[J]. *Nature*, 2019, 556(7743): 195–204.
- Oruganti Y D, Yuan P, Inanc F, et al. Role of Machine Learning in Building Models for Gas Saturation Prediction[C]. Society of Petrophysicists and Well-Log Analysts, 2019.
- Xu C, Misra S, Srinivasan P, Ma S. When Petrophysics Meets Big Data: What can Machine Do? [C]. Society of Petroleum Engineers, 2019.
- Wu H H, Pan L, Ma J, et al. Enhanced Reservoir Geosteering and Geomapping from Refined Models of Ultra-Deep LWD Resistivity Inversions Using Machine-Learning Algorithms[C]. Society of Petrophysicists and Well-Log Analysts, 2019.