



移动扫码阅读

要惠芳,赵明坤,陈 强.基于机器学习的煤系致密砂岩气储层分类研究—以鄂尔多斯盆地 DJ 区块为例[J].
煤炭科学技术,2022,50(6):260-270.
YAO Huifang,ZHAO Mingkun,CHEN Qiang.Research on classification of tight sandstone gas reservoir in coal meas-
ures based on machine learning:a case from DJ Block of Ordos Basin[J].Coal Science and Technology,2022,
50(6):260-270.

基于机器学习的煤系致密砂岩气储层分类研究 ——以鄂尔多斯盆地 DJ 区块为例

要惠芳^{1,2},赵明坤³,陈 强^{1,2}

(1.太原理工大学 矿业工程学院,山西 太原 030024;2.煤与煤系气地质山西省重点实验室,山西 太原 030006;

3.河南省煤炭地质勘察研究总院,河南 郑州 450052)

摘 要:鄂尔多斯盆地 DJ 区块煤系致密砂岩气储层为特低孔隙度和渗透率储层,强非均质性是储层预测和评价的关键难题。利用 1 880 块岩芯柱塞孔渗及 33 组压汞分析资料,计算了流动层段指数和孔喉半径概率累积分布函数,优选了基于流动层段指数(FZI)的流动单元 3 分类方案;为解决不同类型砂岩样本数量失衡的问题,对 734 组宏观岩石相等类别型数据及数值型测井数据进行了二次分层抽样均化采样处理,重构了数据子集;通过参数设定—训练—建模—测试—性能评估的交互验证法优选了建模参数,构建了无监督和有监督 2 类 7 种算法的机器学习模型,分别为 K 均值聚类、朴素贝叶斯、决策树、支持向量机、深度学习、随机森林和梯度提升树;并依据平衡分数 f_1 得分大小确定了梯度提升树学习和深度学习 2 个最优分类模型。研究表明:基于 FZI 流动单元的砂岩分类,能显著降低非均质性对渗透率预测精度的影响;与无监督方法相比,有监督的、积极的机器学习类方法更契合砂岩流动单元分类原则;梯度提升树学习算法能利用岩石相、成岩相、沉积微相等类别型数据建模,可用于流动单元成因特征等基础地质方面的研究,而深度学习对测井等数值型数据处理能力更强,适合于渗透率预测等工程应用;实现致密砂岩流动单元合理分类的机器学习模型并不惟一,可结合数据类型和研究目的优选算法独立构建。

关键词:鄂尔多斯盆地;致密砂岩;非均质性;流动单元;机器学习;渗透率;储层预测

中图分类号:P631 文献标志码:A 文章编号:0253-2336(2022)06-0260-11

Research on classification of tight sandstone gas reservoir in coal measures based on machine learning:a case from DJ Block of Ordos Basin YAO Huifang^{1,2},ZHAO Mingkun³,CHEN Qiang^{1,2}

(1.Department of Earth Science and Engineering,Taiyuan University of Technology,Taiyuan 030024,China;2.Shanxi Key Laboratory of Coal and Coal-measure Gas Geology,Taiyuan 030006,China; 3. Henan Provincial Coal Geological Survey and Research Institute, Zhengzhou 450052, China)

Abstract:The coal-measure tight sandstone in the DJ block of the Ordos Basin is an ultra-low porosity and permeability reservoir, and strong heterogeneity is a key problem in reservoir prediction and evaluation. Based on 1 880 rock cores plug porosity and permeability analysis data and 33 sets of mercury intrusion analysis data, the probability cumulative distribution function of the flow interval index and pore throat radius was calculated, and the three flow units classification scheme with FZI was optimized. Tosolve the problem of the imbalance of the number of different types of sandstone samples, 734 sets of macro-rock equivalent type data and numerical logging data were subjected to sub-stratified sampling and homogenization sampling processing, and the data subset was reconstructed;Through the interactive

收稿日期:2021-10-02 责任编辑:周子博 DOI:10.13199/j.cnki.cst.2020-0791
基金项目:十三五国家重大专项基金资助项目(2016ZX05065);山西省应用基础研究资助项目(201801D121032);阳泉煤业(集团)股份有限公司
资助项目(GY18027)
作者简介:要惠芳(1959—),女,山西汾西人,教授,硕士生导师。E-mail:yhf5908@163.com
通讯作者:赵明坤(1984—),男,河南开封人,高级工程师,博士。Tel:0371-67975202,E-mail:153903939@qq.com

verification method of parameter setting-training-modeling-testing-performance evaluation, the modeling parameters were optimized, and machine learning models of 7 algorithms in two categories, unsupervised and supervised, were constructed. They are K-means clustering, naive Bayes, decision tree, support vector machine, deep learning, random forest and gradient tree boosting respectively, and the two optimal classifications of gradient boosting tree learning and deep learning are determined according to the f_1 score value. The research results show that the sandstone classification based on FZI flow unit can significantly reduce the influence of heterogeneity on the accuracy of permeability prediction; compared with unsupervised method, the supervised and active machine learning methods are more suitable for the classification principle of flow unit. Gradient boosting tree learning algorithm can use lithofacies, diagenetic facies, sedimentary micro-equal type data modeling, and can be used for basic geological research such as the genetic characteristics of flow units, while deep learning has stronger processing capabilities for numerical data such as well logging and is suitable for engineering applications such as permeability prediction. The machine learning model with efficient and reasonable classification for flow units is not unique and can be constructed independently according to research purpose and data type to optimize the algorithm.

Key words: Ordos Basin; tight sandstone reservoir; heterogeneity; flow unit; machine learning; penetration; reservoir prediction

0 引言

致密砂岩气资源量约占到全球非常规资源量的 70%, 是重要的接替能源^[1]; 鄂尔多斯盆地作为我国主要的深部致密砂岩气产区, 其上古生界石炭、二叠系储层是典型的以低、特低孔隙度和渗透率为特征的准连续型气藏, 物性非均质性较强^[2-4]。流动单元, 又称岩石物理流动单元^[5]、水动流动单元^[6-7], 是受沉积、成岩作用等因素控制的、以非均质性为特征的不同类型储集单元, 而在相同类型单元内部岩石物理性质相似, 流体渗流特征相似^[8], 可近似为均质性的或可预测的^[5,9]。因此, 流动单元的合理分类是决定是否可以将低孔低渗储层物性的非均质问题简化为均质问题的关键。流动单元的划分主要有成因特征分类法及微观孔隙结构统计分析法 2 类。前者全部或部分考虑了沉积相、岩石相、成岩和构造作用等地质控制因素, 具体表征参数包括矿物组分和粒度、成熟度及分选性等岩石学特征^[10], 压实、胶结、溶蚀、破裂等成岩作用特征^[6], 高分辨率层序及成因砂体组合、沉积微相及界面, 隔层、夹层、沉积构造等沉积学特征^[11], 断层及其封堵性等构造学特征^[12], 划分方法有层对比及相叠加^[10]、层次分析^[13]、灰色关联^[14]等。但该类方法对地质认识依赖大, 效率低, 可实现性较差。而微观孔隙结构法主要利用孔隙度、渗透率、孔喉半径等测试数据, 通过流动层段指数 (FZI) 和孔喉半径 R_{35} 等表征参数, 基于测井数据采用多元线性回归^[15]、聚类分析^[16]、神经网络^[17-19]、支持向量机等算法^[20-21]等, 划分不同类型的流动单元, 进而在同类单元内, 利用孔隙度与渗透率的良好相关性, 建立相应的渗透率解释模型^[22]。由于后者可以将厘米-米尺度的宏观测井参数与微米尺度的孔喉等微观结构、分类算法等多方面因素有机地结合起来, 而且可将取心

井成果推广至未取心井, 预测性强, 更有利于低孔渗气藏的高效分类和评价。因此, 流动单元统计分类方法是目前应用最广泛的储层分类方法^[23], 也是准确评价低孔渗储层物性参数, 提高预测精度的重要途径。但是以往流动单元统计分类研究还存在一些不足, 如数据集的规模较小, 分类算法及模型单一等。各区块在长期勘探开发中积累了海量的数据, 包括数值型 (孔渗测试数据、压汞、测井等) 和类别型或名称型数据 (岩心观察、薄片、X 射线衍射和扫描电镜等反映的岩石相和成岩相特征) 等。近年来, 基于数据驱动的机器学习技术发展迅速^[24], 为充分利用上述数据, 实现流动单元成因及微观孔隙结构特征的综合分类, 加深对致密砂岩储层的认识, 提供了一个极具潜力的工具。基于此, 研究将利用鄂尔多斯 DJ 区块 1 880 块岩心柱塞孔渗分析及 33 组压汞分析资料, 基于概率累积分布函数曲线变化特征及孔-渗相关性, 优选流动单元分类方案; 在此基础上, 结合 734 组宏观岩石相等类别型及数值型测井数据, 尝试建立无监督、有监督 2 类机器学习模型, 并通过 K 均值聚类、决策树、朴素贝叶斯、随机森林、梯度提升树、支持向量机、深度学习 7 种分类模型的性能评估, 探讨致密砂岩流动单元机器分类学习的可行性。

1 DJ 区块致密砂岩储层特征

研究区位于鄂尔多斯盆地东部晋西挠褶带中南段 (图 1), 整体为一倾向 NW 的单斜构造, 地层走向及主要构造轴向均呈 NNE 向; 构造变形较弱, 仅东部有古驿-窑渠背斜、薛关-峪口断层带等; 主要储集层段位于含煤系地层内, 为一套稳定地台构造背景上的浅海陆棚—三角洲—河流、湖泊充填沉积, 包括本溪组、太原组、山西组和下石盒子组, 埋深为 1 500~2 500 m; 发育有多套煤及煤线, 且富含有机

质的暗色泥岩,泥灰岩等优质烃源岩;上石盒子组发育一套稳定的厚度约 100 m 的湖相泥岩,盖层条件较好。本区块虽然从早白垩世以来持续抬升剥蚀,仍然保留了可观的致密气资源,呈现准连续分布、源外成藏、近源聚集的特征^[3];区内砂岩经历了压实

压溶、不稳定组分溶蚀以及碳酸盐、硅质和黏土矿物胶结等多种成岩作用,现处于晚成岩阶段 B 期^[26]。多类型、多期次、不同强度的成岩作用及其叠加对本区砂岩孔隙、裂隙的形成、保存以及破坏产生了极为复杂的影响。

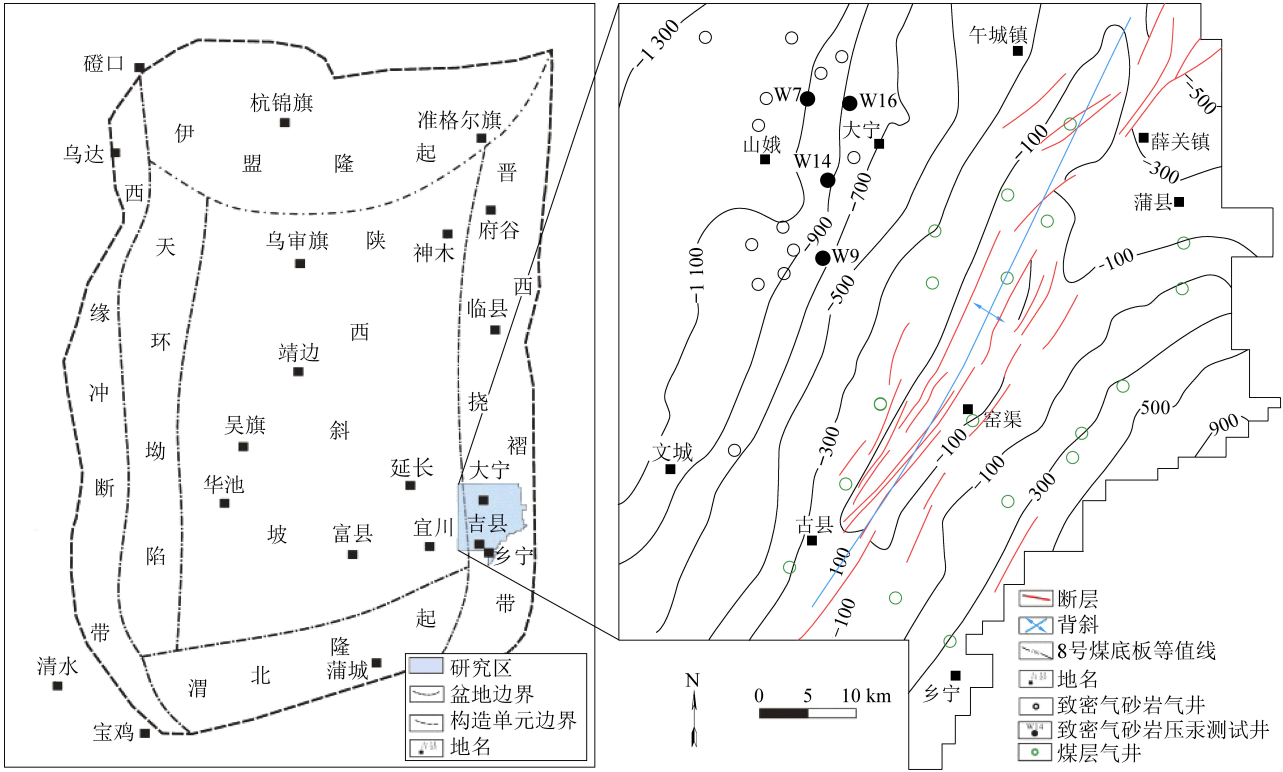


图1 研究区位置及构造纲要图
Fig.1 Location and structure outline map of DJ Block

1.1 致密砂岩储层岩石相特征

已有研究表明^[25],DJ 区块砂岩碎屑成分以石英或岩屑为主,长石含量低,通常不超过 4%;岩屑组分主要为沉积岩,次为变质岩和火成岩。填隙物质量分数为 6%~37%,杂基(以泥质为主)含量一般低于胶结物;在结构特征方面,粒径以中粒居多,粗粒及细粒也有一定的分布;磨圆度以次圆-次棱角状为主,分选好-差;成分成熟度、磨圆度及分选性呈现随地层由老至新逐渐变差的趋势。

根据碎屑组分和粒度,本区砂岩可分 8 种岩石相类型(图 2)。其中,图 2a 为 DJ4-6 井山 2 段粗粒石英砂岩,孔隙较发育,钙质胶结为主,石英次生加大明显;岩屑含燧石、硅质岩、酸性火山岩;图 2b 为 DJ5-8-15 井山 2 段中粒石英砂岩,孔隙发育一般,硅质胶结。岩屑以燧石、硅质岩为主,含少量碳酸盐岩;图 2c 为 DJ4-6 井山 1 段粗粒岩屑石英砂岩,孔隙欠发育,硅质-泥质胶结为主,石英次生加大明显。主要岩屑为变质岩、泥板岩及少量白云母,碳酸盐岩;图 2d 为 DJ6-5 井盒 8 段中粒岩屑石英砂岩,

孔隙发育,粒间与粒内溶蚀孔隙为主,含长石及碳酸盐岩等刚性易溶岩屑;图 2e 为 DJ7 井盒 8 段粗粒岩屑砂岩,孔隙较发育,泥质胶结为主,局部石英有次生加大;岩屑为变质石英岩、碳酸盐岩、燧石、酸性火山岩。图 2f 为 DJ7 井盒 8 段中粒岩屑砂岩,泥质胶结,孔隙较发育;岩屑包括泥板岩、燧石、变质石英岩、碳酸盐岩、酸性火山岩;图 2g 为 DJ5-8 井盒 8 段细粒岩屑石英砂岩,孔隙较发育,硅质胶结为主;岩屑以硅质岩、燧石、酸性火山岩、碳酸盐岩为主;图 2h 为 DJ6-8 井盒 8 段细粒岩屑砂岩,孔隙欠发育,有较弱的粒间与粒内溶蚀孔隙,泥质胶结,溶蚀作用较弱;岩屑主要有泥板岩、片岩、白云母等软岩。

1.2 微观孔、渗特征

根据岩心观察、薄片和 X 射线衍射、扫描电镜结果,本区砂岩孔隙类型主要为原生残余粒间孔及泥质杂基微孔、溶蚀粒间孔和粒内孔、铸模孔、晶间孔和微裂隙等次生孔隙;岩心孔渗资料统计分析表明,区内砂岩孔隙度为 0.60%~19.88%,中值 5.85%,渗透率 $(0.000\ 4\sim 244)\times 10^{-3}\mu\text{m}^2$,中值 0.131×10^{-3}

μm^2 ;DJ 区块虽然以特低孔渗储层为主,但部分层

段发育有大孔喉及微裂缝,渗透率有所改善。

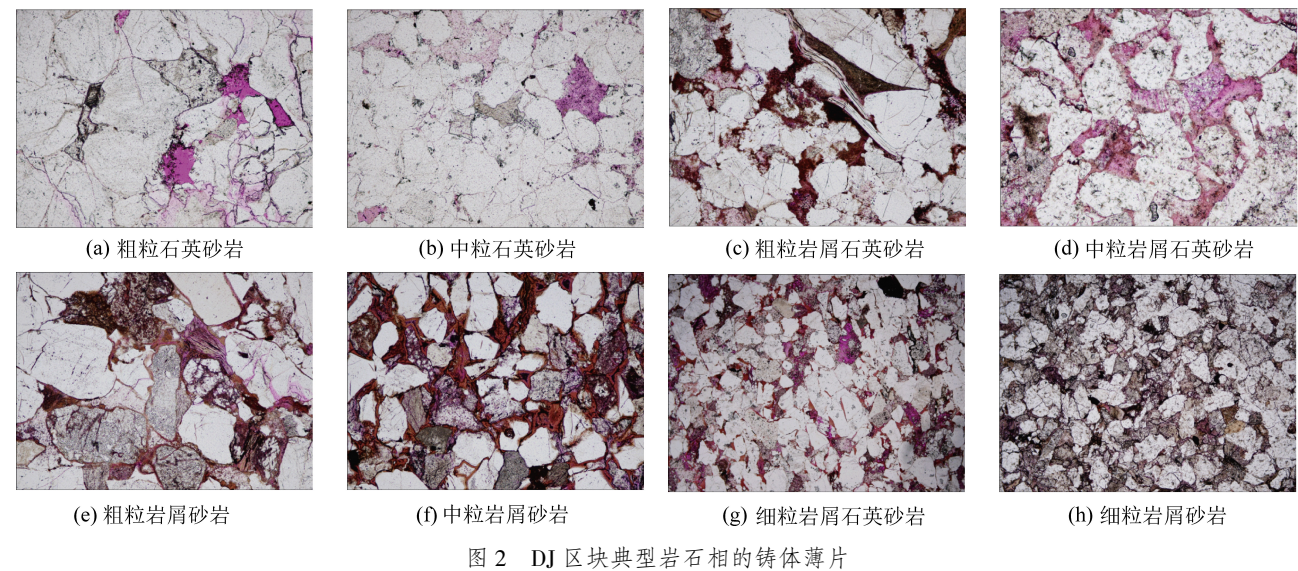


图 2 DJ 区块典型岩石相的铸体薄片

DJ7 等关键井的 33 个岩心柱塞压汞数据表明,储集岩中值压力为 1.0~5.5 MPa,平均值 4.4 MPa,喉道系数 5.272~12.180,中值为 9.35;喉道分选系数 1.414~5.109,中值 2.56;结构系数 0.122~108.100,中值 0.28;偏度 0.028~2.530,平均值 1.51;平均孔喉半径为 0.076~11.115 μm ,中值 0.28 μm ; R_{35} 孔喉半径为 0.016~10.036 μm ,中值为 0.2 μm ,具有显著的小孔径、窄喉道的孔隙结构特征。根据上述特征可

划分 3 种类型(图 3a),第 I 类峰值孔径为 8.0~11.9 μm ,第 II 类 0.12~1.26 μm ,第 III 类 0.012~0.032 μm ;相对应的 35%进汞饱和度毛管压力约为 0.9 MPa(图 3b),20 MPa(图 3c)和 80 MPa(图 3d)。

整体来看,主要受宏观岩石相、沉积相、成岩相等多种地质因素的控制,本区砂岩储集层的孔渗特征非均质性极强。

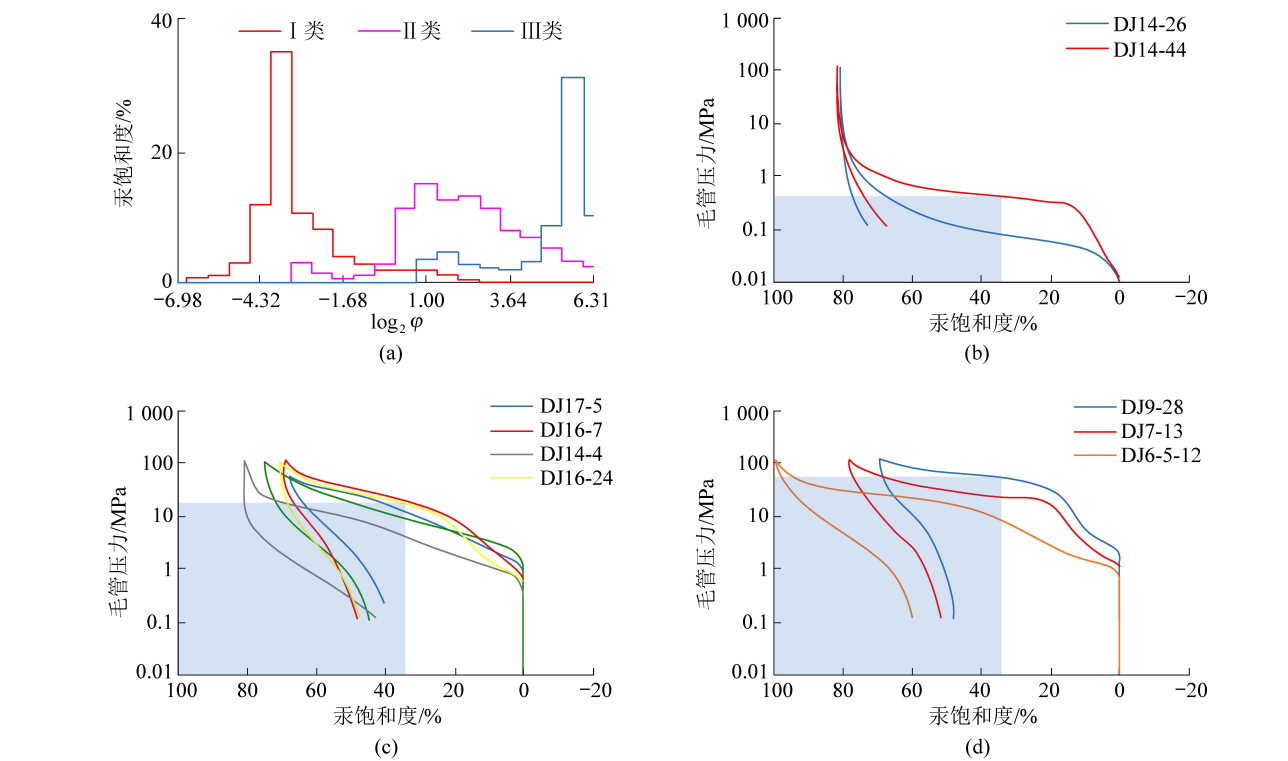


图 3 典型孔径特征分布及其压汞曲线

Fig.3 Distribution characteristics chart of three typical throat size and mercury injection capillary pressure curves

2 流动单元分类

2.1 流动单元划分方法及标准

AMAEFULE 等^[7]提出了采用孔隙度和渗透率实现流动单元分类的 FZI 法,FZI 值计算方法如式(1)一式(3)所示:

$$RQI = 0.314 \sqrt{\frac{k}{\varphi_e}} \tag{1}$$

$$\varphi_z = \frac{\varphi_e}{1 - \varphi_e} \tag{2}$$

$$FZI = RQI / \varphi_z \tag{3}$$

式中:RQI 为储层品质指数,μm;FZI 为流动层段指数; k 为渗透率, $10^{-3} \mu m^2$; φ_e 为有效孔隙度,%; φ_z 为标准化孔隙度,%。

此外,PITTMAN 等^[26]认为压汞曲线上进汞饱和度达 35%时的孔隙半径 R_{35} 能够反映流体流动状

态,因此可用其来划分流动单元;利用孔隙度和渗透率计算孔喉半径 R_{35} 值的方法见式(4)^[27],即

$$\lg R_{35} = 0.523 \lg \varphi - 0.565 \lg k - 0.255 \tag{4}$$

式中: φ 为孔隙度,%。

为更好地反映研究区实际孔渗特征,将 33 组 R_{35} 与孔隙度、渗透率实测值作了二元非线性回归,回归方程见式(5), R^2 可达 0.96。

$$\lg R_{35} = 0.650 \lg k - 1.019 \lg \varphi + 0.105 \tag{5}$$

因此,利用式(3)和式(5)对研究区 1 880 组岩心数据分别计算了 FZI 和 R_{35} ,在此基础上,进一步计算了 $\lg FZI$ 和 R_{35} 的概率累积分布函数(Cumulative density Function)(图 4)。

按照文献[16]中的方法,基于 $\lg FZI$ 和 R_{35} 概率累积分布函数曲线的斜率变化特征,划分了 3 类流动单元,如图 4 和表 1 所示。

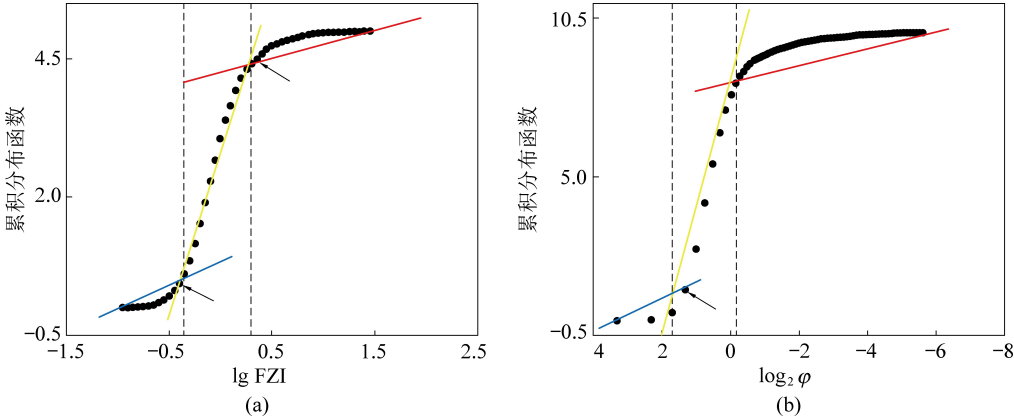


图 4 $\lg FZI$ 及 R_{35} 指数概率累积分布函数散点图
Fig.4 Cumulative density function plots to obtain optimal number of $\lg FZI$ and R_{35} flow units

图 4 中黑色箭头所指的位置即累积密度分布函数曲线斜率的突变点。无论是 $\lg FZI$ 还是 R_{35} 法,除第Ⅱ类单元外,其他 2 类单元内部还可见小的斜率变化,仍可细分;但由于在所有样本中,第Ⅱ类单元中的样本占比(>50%)远大于其他 2 类(<16%),如果再将Ⅰ类和Ⅲ类细分,其样本数量会更少。此外,在后续机器学习前还需做数据清洗工作,数据集之中的样本数还将进一步减少。因此,本次对Ⅰ,Ⅲ类单元划分标准做了部分粗化处理。

表 1 DJ 区块储层流动单元分类

Table 1 Classification scheme of reservoir flow units				
流动单元	流动单元指数		$R_{35}/\mu m$	
类型	划分标准	样本占比/%	划分标准	样本占比/%
Ⅰ类	$\lg FZI \geq 0.2$	14.4	$R_{35} \geq 1.2$	12.3
Ⅱ类	$-0.35 \leq \lg FZI < 0.2$	69.8	$0.4 \leq R_{35} < 1.2$	54.1
Ⅲ类	$\lg FZI < -0.35$	15.9	$R_{35} < 0.4$	33.5

2.2 流动单元分类方案优选

根据表 2 中流动单元划分标准,将 3 类单元中的所有样本投影到 φ - $\lg k$ 坐标系,如图 5 和图 6 所示。渗透率的孔隙度一元回归模型见表 2。通过对比可以发现:① 各类流动单元内孔隙度与渗透率的相关性显著提升,仅有一个单元的 R^2 小于 0.5,其余为 0.68~0.81(表 2)。说明 FZI 和 R_{35} 两种流动单元分类法均能反映微观孔隙结构的变化,且同一流动单元内的样本具有相近的孔喉特征;② 在同一单元内,随着孔隙体积比的增加,渗透率呈幂指数增大;③ 相同孔隙体积条件下,FZI 和 R_{35} 值越大,孔喉半径和渗透率越大,具有Ⅰ类>Ⅱ类>Ⅲ类单元的变化规律;④ FZI 和 R_{35} 两种分类方案的主要差别在于第Ⅰ类单元, R_{35} 划分结果中孔渗相关度仅为 0.47,相对于 FZI 的 0.81 明显变差。因此,本次流动单元分类最终采用了基于流动层段

指数 FZI 的划分方案。

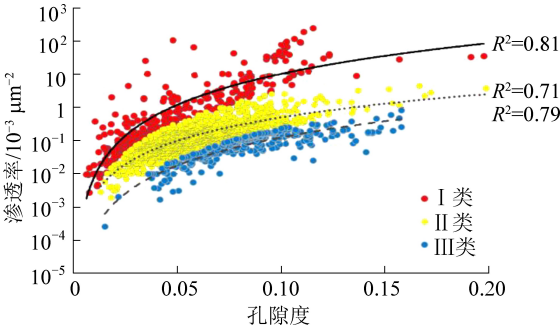


图 5 lg FZI 值划分的流动单元孔、渗分布特征
Fig.5 Three flow units identified by lg FZI indicator in relation to porosity and permeability

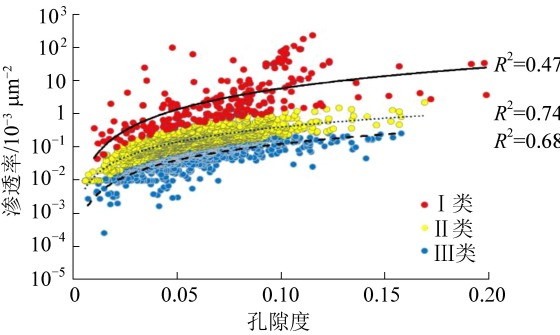


图 6 R_{35} 值划分的 3 类流动单元的孔、渗分布特征
Fig.6 Three flow units identified by R_{35} indicator in relation to porosity and permeability

表 2 DJ 区块 3 类流动单元渗透率解释模型
Table 2 Permeability interpretation models of 3 flow units in DJ Block

流动单元 类型	流动单元指数		R_{35} 指数/ μm	
	渗透率的孔隙度 回归解释模型	相关 系数 R^2	渗透率的孔隙度 回归解释模型	相关 系数 R^2
I 类	$11\ 940\varphi^{3.071\ 3}$	0.81	$0.044\ 6\varphi^{2.12}$	0.47
II 类	$108.82\varphi^{2.336\ 3}$	0.71	$0.012\ 2\varphi^{1.52}$	0.74
III 类	$76.03\varphi^{2.788\ 7}$	0.79	$0.002\ 9\varphi^{3.071\ 3}$	0.68

综上,合理的流动单元划分方案可以改善微观孔隙结构的均质性问题,提高孔隙度和渗透率的相关度。这样,当砂岩样所属流体单元判别正确时,渗透率的预测精度将会大幅提高。

3 机器学习分类建模及效果

鉴于大多数井不取心,仅有测井和录井资料。因此,如何以测井、录井数据为媒介,将岩心测试成果统计分析确定的流动单元的分类成果,推广至未取心井,实现流动单元的高精度分类,是本次流动单元分类机器学习重点解决的问题。

3.1 机器学习数据集及数据处理

3.1.1 初始数据集

共收集到 46 井 1 880 组的岩心孔隙度、渗透率测试数据以及少量的岩心观察、薄片、岩电测试分析等资料;但相对应且有完整测、录井信息的样本较少,最终可用于机器学习的仅有 24 井 808 个样本。

3.1.2 数据属性

测井选取了井径(CAL)、补偿中子(CNL)、自然伽玛(GR)、自然电位(SP)、声波时差(AC)、补偿密度(DEN)、视电阻率(R_a , 8 inch 感应电阻率)、岩性密度(PE)、测井孔隙度(POR)等 9 种参数属性;此外,通过计算样点埋深以上岩石的垂向静压力构建了表征应力差异的垂向应力属性;利用各层段沉积的地质年代(时间)构建了年代属性,具体属性值有盒 7,盒 8,山 1,山 2,太原,本溪共 6 个;利用录井资料构建了岩石相属性,具体属性值有粗粒石英砂岩、中粒石英砂岩、粗粒岩屑石英砂岩、中粒岩屑石英砂岩、粗粒岩屑砂岩、中粒岩屑砂岩、细粒岩屑砂岩等 7 个。以上 12 个属性中,岩石相、年代属性为类别型数据,其他均为数值型。

3.1.3 分类标签

按照表 1 中 FZI 指数的分类标准,将 lg FZI 的 2 个特征值-0.35 和 0.2 作为门槛,定义了 3 类流动单元,并对所有样本设定了相应的分类标签,作为输出变量,用于有监督学习的训练和测试。

3.1.4 数据集处理

通过深度校正、环境校正和标准化等预处理,使测井数据在区内具有统一的刻度,并使系统误差减至最小。对电阻率做了对数处理,使其数据分布呈正态化;自然电位进行了泥岩基线校正;以井径增大率小于 50%为门槛对数据进行了筛选,筛选后总样本数减少至 24 井 734 组。此外,针对部分异常值和噪音做了零均方差(Z-score)标准化处理;针对不同属性值数据范围差异大的问题,基于不同算法要求在学习过程中做了线性或非线性归一化处理。

3.2 分类算法

机器学习模式可分为 2 类:无监督与有监督学习。研究的重心是测试有监督学习算法,以期对测井、岩石相等数据(输入)和流动单元类型(输出)之间的关系进行泛化和评估,最终实现分类模型优选。因此,只采用 K 均值聚类算法进行了无监督学习试验,而测试的有监督学习算法共 5 种,包括朴素贝叶斯、决策树、支持向量机、深度学习和集成学习算法(随机森林、梯度提升树)。6 种机器学习算法原理简述如下:① K 均值聚类(K-means Clustering,

KC),是一种无监督学习方法,需要给定簇的个数,找出各簇的质心,将与各质心距离相近的数据点聚集成簇,实现分类;② 朴素贝叶斯(Naive Bayesian, NB),是一种统计学方法,通过对相互独立的多个输入变量,应用贝叶斯理论计算各个类别的概率,选取最大概率对应的类别作为分类结果;③ 决策树(Decision Trees, DT),是一种信息论方法,通过信息熵构建纯的子集,对目标变量影响最显著的将在树的根或近根层次;④ 支持向量机(Support Vector Machine, SVM),采用数学最优化理论,通过构建多维的超平面形成多个界面进行分类;⑤ 深度学习(Deep Learning, DL),通过模拟神经元的工作方式,建立具多隐层结构且有反馈的网络来连接输入和输出,从而实现分类;⑥ 集成学习,一种“元学习”方法,通常由多个不同的独立模型,利用集体智慧如投票的方式,综合成一个最优的模型。该类算法非常适合噪声数据、离群点存在的数据集,本次测试的随机森林(Random Forest, RF)和梯度提升树(Gradient Tree Boosting)均为以决策树为基础模型的集成学习类算法。

本次机器学习分类的所有算法均在 Rapidminer Studio 数据挖掘平台上实现。

3.3 训练及测试数据集

通过以上方式处理后的数据集中共包含 734 组样本(24 井),其中 I 类样本 111 个, II 类 496 个, III 类 127 个;各分类样本数失衡,为提高后续建模的可信度,须作均化处理^[24]。为保证有相同的数据分布特征,按二次分层抽样(Stratified Sampling)方式对 II 类数据进行了 1/3 比例抽稀,并分别与 I 类和 III 类样本合并,形成 3 个数据子集分别进行建模。每个子集内 3 类样本数量一致,均为 111-165-127,总样本数量为 503。每个子集均按 50%~90%的比例抽样训练,剩余样本作为测试集进行验证和评估;最终的模型性能指标以 3 个子集的等权重之和计算,再进行评估。

针对不同的算法特性,本次机器学习的训练和测试选择了不同的属性集合:K 均值聚类、深度学习、支持向量机等全部采用纯数值型数据,具体包括 CNL、GR、SP、AC、DEN、Ra、PE、垂向应力等 8 种;朴素贝叶斯、决策树、随机森林和梯度提升树采用数值型和类别型的混合型数据;包括年代、岩石相、POR、CNL、GR、SP、Ra、垂向应力等 8 种。

3.4 建模参数优选

每种算法均需要与特定数据集与方法本身匹配的建模参数,如决策树的树个数、支持向量机的学习

速率等。上述参数如用人工进行优选,效率低,客观性难以保证。为此,采用参数区间及其步长设定-训练-建模-测试-性能评估的交互验证方法,其优点在于,每一组参数的改变均可由计算机自动实现训练、建模和测试,并获得其所建模型的分类精度、召回率和平衡分数 f_1 值等。由于 f_1 能综合反映分类精度和召回率的变化^[24],相对更全面,因此,可以通过 f_1 箱形图的变化特征对各算法的建模参数及其组合进行分析和优选,以保证各算法均能实现最佳分类。

不同算法对于建模参数的敏感度不同, f_1 离散度越高就越敏感^[28]。图 7 为 6 种有监督学习算法的不同建模参数组合 f_1 得分值箱形图,在图 7 中,朴素贝叶斯 f_1 分布最为集中,深度学习和支持向量机 f_1 离散度最高,以决策树为基础的 3 种算法介于两者之间,说明深度学习和支持向量机 2 种算法对建模参数要求是最高的;此外,二者虽然最高和最低得分基本相同,但 f_1 的中位数差异较大,深度学习为 78.8%,而支持向量机仅 73.5%,说明后者只有少数参数组合有较好的分类表现,算法对参数依赖性强,稳定性较差。

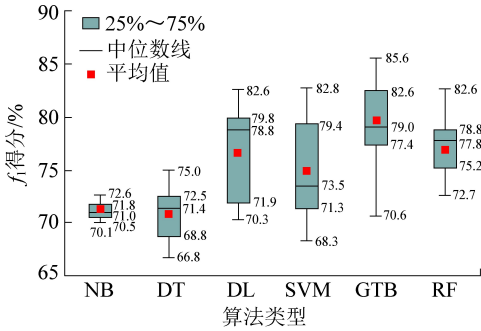


图 7 不同算法建模参数组合的验证 f_1 得分箱形图

Fig.7 Boxplots of f_1 -score obtained from cross validation performed over different parameter sets of each algorithm

通过上述方法,可以得到每种算法的最优参数组合。以梯度提升树模型为例,其最优参数为:树个数:100;可再生性:真实;本地随机种子数:1992;最大深度:10;学习率:0.02;分布形式:多模态。

3.5 模型性能评估及模型优选

3.5.1 有监督与无监督学习模型

图 8 为无监督学习 K 均值聚类方法与其他 5 种有监督学习的最终分类效果对比图,与图 5 相同,红色样点代表 I 类流动单元砂岩,黄色为 II 类,蓝色为 III 类。可见,有或无监督学习这 2 类机器学习模型的样点分布特征大相径庭,5 种有监督学习算法的样点分布与图 5 相近,而 K 均值聚类模型结果与

FZI 分类基本不相关。这一现象说明同样基于数据驱动,以消极学习为特征的 K 均值聚类只有在数据本身与分类目标期望契合的特定情况下,才可能实现理想分类;而以深度学习为代表的积极学习算法,在类标签的约束下,对隐含规则或关系的泛化能力

更强,知识挖掘能力更高。因此,如若应用无监督的机器学习算法对流动单元进行分类,训练测试数据集的构建至关重要;而有监督的、积极的学习方法不需要对数据集进行针对性处理,学习门槛低,效率更高。

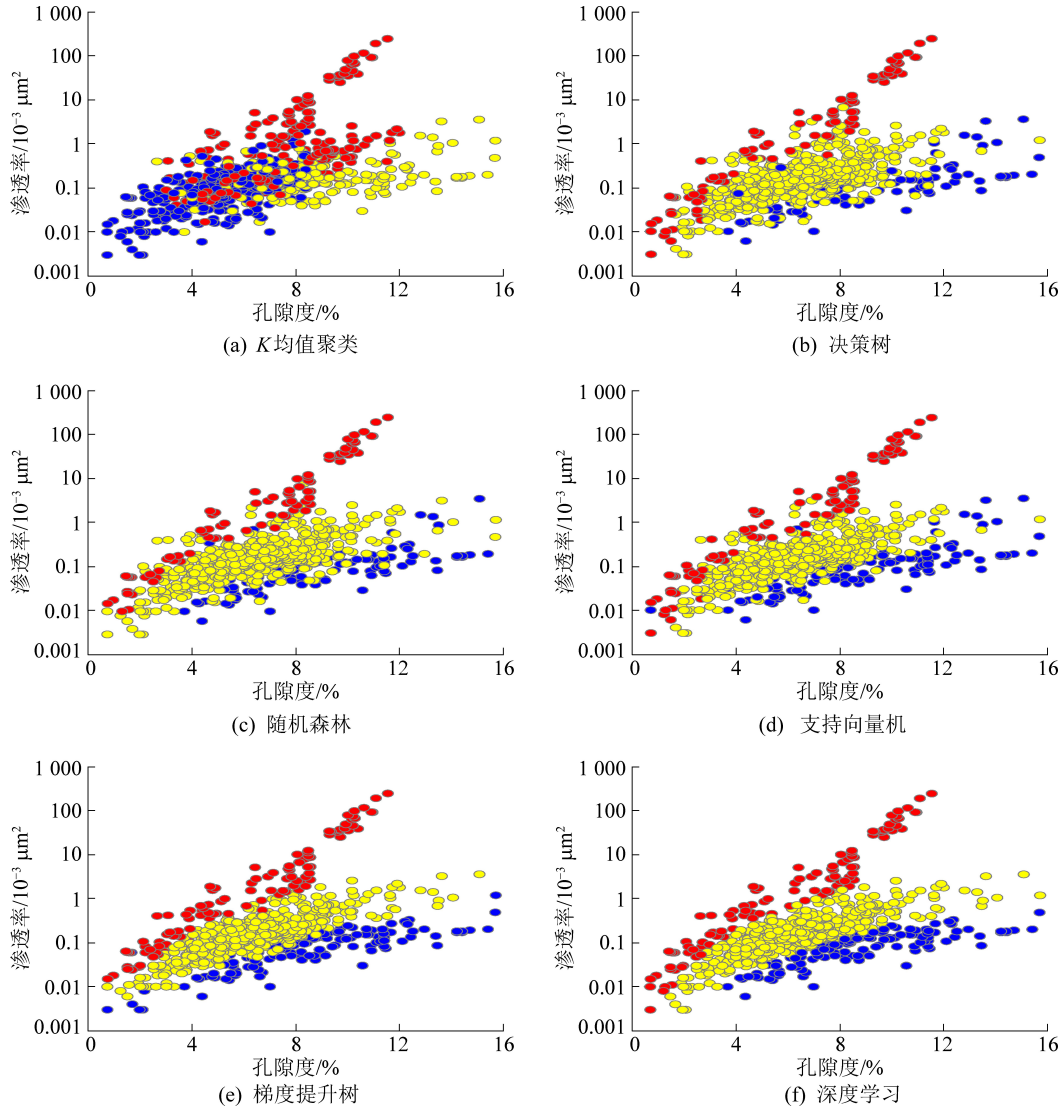


图 8 不同机器学习算法分类结果对比

Fig.8 Classification results of different machine learning algorithms

3.5.2 有监督学习模型优选

该类模型的整体表现较好, f_1 最高值均超过了 72% (图 7),可以对流动单元进行有效分类。另外,从中值分布特征来看,朴素贝叶斯最低仅为 71%;深度学习、梯度提升树和随机森林 2 类集成学习算法的表现较好,中值分别为 78.8%、79.0%和 77.8%,接近 80%,而朴素贝叶斯、决策树和支持向量机较低,为 70%~74%。

综合 f_1 分布特征 (图 7) 和分类效果 (图 8),本次机器学习最终确定的最优模型有 2 个:一个是基于数值型数据集的深度学习模型,另一个是面向数

值和类别型混合数据的梯度提升树模型,其 f_1 平均得分值分别为 82.63%和 85.56%,其性能指标详见表 3。

可见,在 DJ 区块,采用不同的数据类型和与之匹配的机器学习算法,可以构建出准确度和可靠性相近的不同机器学习模型。因此,能够实现流动单元高效合理分类的机器学习模型并不是唯一的。

对于工程应用来说,可以基于井径、补偿中子、自然伽玛、自然电位、声波时差、补偿密度、视电阻率、岩性密度等表征岩性、物性和流体性质的测井参数,通过构建以深度学习为代表的数值型机器学习

表3 梯度提升树和深度学习分类模型性能对比

Table 3 Classification matrices obtained with prediction over training set using GTB and DL algorithms

分类 算法	流动单 元类型	精确率/ %	召回率/ %	f_1 得分	f_1 平均 得分
梯度提升树	I 类	86.05	77.17	79.61	85.56
	II 类	94.40	91.94	93.15	
	III 类	82.22	81.90	83.92	
深度学习	I 类	86.84	78.63	82.53	82.63
	II 类	90.42	89.65	90.03	
	III 类	85.18	67.52	75.32	

模型,实现流体单元的准确分类,进而提高渗透率的测井解释精度。虽然这一类模型难以解读其地质含义,但是方法健壮性好,分类预测精度高,既能处理高度非线性问题,又有极强的反馈能力^[28],是高效完成现场准确预测任务的最佳选择。

但对于流体单元的成因特征等基础地质研究来说,由于其采用的地质语言均为类别型,如不同类型砂岩的形成年代、岩石相、沉积微相、成岩相等,很难转换为数值型数据加以量化表征,深度学习、支持向

量机等数值类机器学习算法无法处理。能够对上述信息进行处理只有朴素贝叶斯、决策树和以其为基础的梯度提升树、随机森林等集成学习模型,而且模型结果兼具可解读性^[29]。虽然本次研究仅考虑了岩石相和年代特征,但是 DJ 区块梯度提升树学习模型研究初步表明(图 9):排在前 13 位的建模影响因子中(影响因子总和为 81.8%),测井孔隙度等 6 种数值型属性贡献率之和为 45.1%,粗粒石英砂岩等 7 种岩石相类别属性为 36.7%,说明二者影响力相当,岩石相同样对最终模型的构建有显著影响。同时也可以看到,岩石相的属性变化对本区砂岩储层渗流特性的影响程度明显高于年代属性(未进入前 13 位),控制作用更明显。如果能进一步考虑成岩相、沉积微相等其他类似岩石相的地质信息,建立更完善的数据集,利用已有的 FZI 流体单元分类结果,完成训练、建模和测试,就能通过解读最优模型泛化产生的具体分类指标,挖掘其关联关系和分类规则,归纳分析不同相及其耦合对流体单元分类的影响程度,从而形成关于研究区砂岩渗流特性成因的地质控制要素及其耦合关系的新认识。

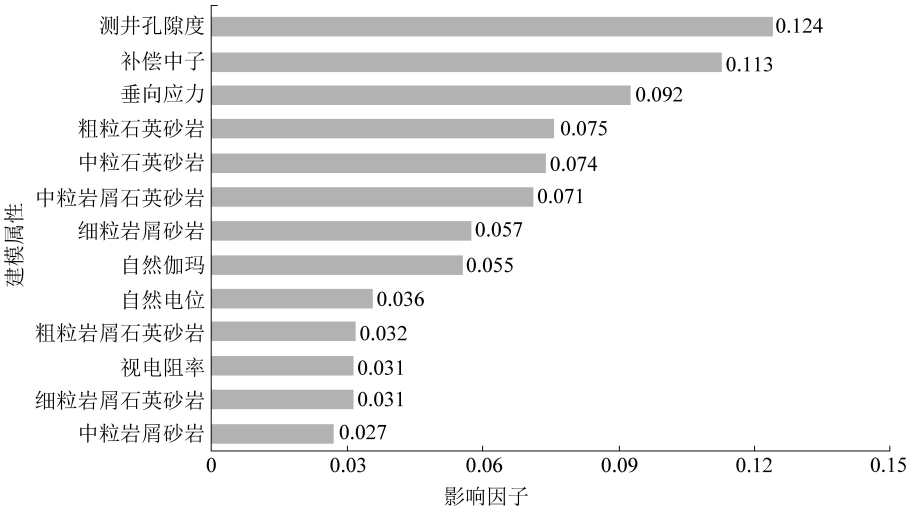


图9 梯度提升树模型部分建模属性影响因子

Fig.9 Relative importance of input features for classification with gradient tree boosting algorithm

因此,采用不同的数据集和与之相适应的机器学习算法,可以构建相互独立的流体单元分类模型,进而实现不同的研究目标。

4 结 论

1)利用 1 880 块岩芯柱塞孔渗及 33 组压汞数据,通过 lg FZI 和 R_{35} 概率累积分布函数曲线和渗透率、孔隙度一元回归分析,优选了 DJ 区块致密砂岩储层的 FZI 流动单元 3 分类方案。在同一单元内,渗透率与孔隙度相关性增强,呈幂指数关系,有利于

提高渗透率的定量预测精度。

2)针对 II 类流动单元砂岩样本数量远大于 I、III 类砂岩,数据分布失衡的问题,采用二次分层抽样方法构建了不同类型砂岩样本数量相近的多个数据子集,增强了有监督机器学习模型的可靠性。

3) 综合应用交互验证和 f_1 得分箱形图法实现了不同机器学习算法的建模参数优选和算法稳定性分析。朴素贝叶斯对参数选取不敏感,决策树、随机森林和梯度提升树次之,而深度学习和支持向量机等 2 种算法对建模参数及其组合最为敏感,稳定性

相对较差。

4) 与无监督方法相比,有监督的、积极的机器学习方法更适合于流动单元分类研究;在 DJ 区块所有砂岩机器学习分类模型中,深度学习和梯度提升树算法模型表现最优。

5) 深度学习模型对测井参数等数值型数据的处理能力强,适用于渗透率定量解释等工程应用场景;梯度提升树等以决策树为基础的集成建模方法可以处理岩石相、成岩相、沉积微相等有明确地质含义的类别型数据,模型的可解读性优于深度学习,在流体单元成因地质研究方面有较好的应用潜力。

参考文献 (References) :

[1] 李五忠,孙 斌,孙钦平,等.以煤系天然气开发促进中国煤层气发展的对策分析[J].煤炭学报,2016,41(1):67-71.
LI Wuzhong,SUN Bin,SUN Qinqing, *et al.* Analysis on coal-bed methane development based on coal measure gas in China and its countermeasure [J]. Journal of China Coal Society,2016,41(1):67-71.

[2] 高丽军,逢建东,谢英刚,等.临兴区块深部煤层气潜在可采地质模式分析[J].煤炭科学技术,2019,47(9):89-96.
GAO Lijun,PANG Jiandong,XIE Yinggang, *et al.* Analysis on potential geological mining model of deep coalbed methane in Linxing Block[J]. Coal Science and Technology,2019,47(9):89-96.

[3] 谢英刚,秦 勇,叶建平,等.临兴地区上古生界煤系致密砂岩气成藏条件分析[J].煤炭学报,2016,41(1):181-191.
XIE Yinggang,QIN Yong,YE Jianping, *et al.* Accumulation conditions of tight gas in the Upper Paleozoic of Linxing Block[J]. Journal of China Coal Society,2016,41(1):181-191.

[4] 牛鑫磊,曹代勇,徐 浩,等.海陆过渡相煤系致密砂岩储层特征及影响因素[J].煤炭科学技术,2018,46(4):188-195.
NIU Xinlei,CAO Daiyong,XU Hao, *et al.* Characteristics and control factors of tight sandstone reservoirs in marine-continental transitional coal measures [J]. Coal Science and Technology,2018,46(4):188-195.

[5] EBANKS W J Jr. Flow unit concept; Integrated approach to reservoir description for engineering projects [J]. AAPG Bulletin,1987,71(5):551-552.

[6] 李波波,王 斌,杨 康,等.煤岩孔裂隙结构分形特征及渗透率模型研究[J].煤炭科学技术,2021,49(2):226-231.
LI Bobo,WANG Bin,YANG Kang, *et al.* Study on fractal characteristics of coal and rock pore fissure structure and permeability model[J]. Coal Science and Technology,2021,49(2):226-231.

[7] AMAEFULE J O,ALTUNBAY M,TIAB D, *et al.* Enhanced reservoir description; using core and log data to identify hydraulic (flow) units and predict permeability in un-cored interval/ wells [J]. SPE 26436,1993,205-220.

[8] 袁彩萍,姚光庆,徐思煌,等.油气储层流动单元研究综述[J].地质科技情报,2006,25(4):23-28.
YUAN Caiping,YAO Guangqing,XU Sihuang, *et al.* Review on fluid flow unit in oiland gas reservoirs [J]. Geological Science and

Technology Information,2006,25(4):23-28.

[9] 冯晓宏,刘学峰,岳清山,等.厚油层非均质特征描述的新方法:水力(渗流)单元分析[J].石油学报,1994,15(S1):149-158.
FENG Xiaohong,LIU Xiufeng,YUE Qingshan, *et al.* A new method for heterogeneity description of thick reservoir; hydraulic flow units analysis [J]. Acta Petrolei Sinca,1994,15(S1):149-158.

[10] 熊琦华,彭仕宓,黄述旺,等.岩石物理相研究方法初探:以辽河凹陷冷东:雷家地区为例[J].石油学报,1994,15(S1):68-75.
XIONG Qihua,PENG Shimi,HUANG Shuwang, *et al.* A preliminary study of the new concept of petrophysical facies and its initial application in Lengdong-Leijia region in Liaohe depression [J]. Acta Petrolei Sinca,1994,15(S1):68-75.

[11] HAMLIM H S,DUTTON S P,SEGIE D, *et al.* Depositional controls on reservoir properties in a Braid - Delta sandstone, Tirrawarra Oil Field. South Australia [J]. AAPG Bulletin,1996,80(2):139-156.

[12] LARUE D K,LEGARRE H. Flow units, connectivity and reservoir characterization in a wave-dominated deltaic reservoir: meren reservoir, nigeria [J]. AAPG Bulletin,2004,88(3):303-324.

[13] 焦养泉,李思田,李 祯,等.碎屑岩储层物性非均质性的层次结构[J].石油与天然气地质,1998,19(2):89-92.
JIAO Yangquan,LI Sitian,LI Zhen, *et al.* Heterogeneity of porosity and permeability in clastic rock reservoirs [J]. Oil & Gas Geology,1998,19(2):89-92.

[14] 宋子齐,杨立雷,王 宏,等.灰色系统储层流动单元综合评价方法[J].大庆石油地质与开发,2007,26(3):76-81.
SONG Zhiqi,YANG Lilei,WANG Hong, *et al.* Comprehensive evaluation method of reservoir flow unit with grey system [J]. Petroleum Geology and Oilfield Development in Daqing,2007,26(3):76-81.

[15] 范宜仁,葛新民,汪海龙,等.非均质性砂砾岩储层渗透率预测方法研究[J].西南石油大学学报,2010,32(3):6-10.
FAN Yiren,GE Xinmin,WANG Hailong, *et al.* Study on the method predicting permeability in the heterogeneous glutenite reservoir [J]. Journal of Southwest Petroleum University,2010,32(3):6-10.

[16] KADKHODAIE I R,RAHIM R,REZA H H, *et al.* Analysis of the reservoir electro-facies in the framework of hydraulic flow units in the Whicher Range Field, Perth Basin, Western Australia [J]. Journal of Petroleum Science and Engineering,2013,111:106-120.

[17] AMINIAN K,AMERI S,OYEEROKUN A, *et al.* Prediction of flow units and permeability using artificial neural networks [J]. Eastern Mediteranean Health Journal,2003,16(6):615.

[18] 司马立强,肖 华,袁 龙,等.基于 BP 神经网络技术的储层流动单元研究[J].测井技术,2012,36(4):421-425
SIMA Liqiang,XIAO Hua,YUAN Long, *et al.* Research on reservoir flow unit based on BP neural network technology [J]. Well Logging Technology,2012,36(4):421-425.

[19] 陈志强,吴思源,白 蓉,等.基于流动单元的致密砂岩气储层

渗透率测井评价:以川中广安地区须家河组为例[J].岩性油气藏,2017,29(6):76-83.

CHEN Zhiqiang, WU Siyuan, BAI Rong, *et al.* Logging evaluation for permeability of tight sandstone gas reservoirs based on flow unit classification: a case from Xujiache formation in Guang'an area, central Sichuan Basin [J]. Lithologic Reservoirs, 2017, 29(6):76-83.

[20] 杨爱东,单立群,刘彦昌,等.基于WT与LSSVM的储层流动单元划分方法[J].测井技术,2017,41(2):237-242.

YANG Aidong, SHAN Liqun, LIU Yanchang, *et al.* Method of flow unit identification for reservoir formation based on WT and LSSVM[J]. Well Logging Technology, 2017, 41(2):237-242.

[21] 袁伟,张占松,张泽宇,等.基于储层分类的支持向量机渗透率预测[J].测井技术,2015,39(4):511-514.

YUAN Wei, ZHANG Zhansong, ZHANG Zeyu, *et al.* Permeability prediction using support vector machine based on reservoir classification[J]. Well Logging Technology, 2015, 39(4):511-514.

[22] 王月莲,宋新民.按流动单元建立测井储集层解释模型[J].石油勘探与开发,2002,29(3):53-55.

WANG Yuelian, SONG Xinmin. New method for well log interpretation by single fluid flow unit[J]. Petroleum Exploration and Development, 2002, 29(3):53-55.

[23] 吕明针,林承焰,张宪国,等.储层流动单元划分方法评价及优选[J].岩性油气藏,2015,27(1):74-80.

LYU Mingzhen, LIN Chengyan, ZHANG Xianguo, *et al.* Evaluation and optimization of flow unit division methods[J]. Lithologic Reservoirs, 2015, 27(1):74-80.

[24] CATÉ A, L P ERWAN G, MARTIN B. Machine learning as a tool for geologists[J]. The Leading Edge, 2017, 36(3):215-219.

[25] 姚海鹏,李玲,周晓刚.鄂尔多斯盆地上古生界煤系非常规天然气赋存特征[J].煤炭科学技术,2017,45(4):102-109,136.

YAO Haipeng, LI Ling, ZHOU Xiaogang. Deposition features of unconventional natural gas in Upper Paleozoic coal measures of Ordos Basin[J]. Coal Science and Technology, 2017, 45(4):102-109,136.

[26] PITTMAN E D. Relationship of porosity and permeability to various parameters derived from mercury injection-capillary pressure curves for sandstone[J]. AAPG Bulletin, 1992, 76(2), 191-198.

[27] ELSHAN A, MILAD S, MANIKA P, *et al.* Rock typing of tight gas sands: a case study in Lance and Mesaverde formations from Jonah field[J]. Journal of Natural Gas Science and Engineering, 2016, 33:1260-1270.

[28] VIJAY K Bala D. Predictive analytics and data mining: concepts and practice with RapidMiner[M]. Singapore: Posts & Telecom Press, 2015.