沉积学报
ACTA SEDIMENTOLOGICA SINICA
ISSN 1000-0550
CN 62-1038/P

机器学习方法在浅层滩坝相薄储层孔隙度预测中的应用

张宇航,时保宏,张曰静,石好果,文雯,张杨

第41卷 第5期 沉积学报 Vol.41 No.5
2023年10月 ACTA SEDIMENTOLOGICA SINICA Oct.2023

文章编号:1000-0550(2023)05-1559-09 DOI: 10.14027/j.issn.1000-0550.2022.100

机器学习方法在浅层滩坝相薄储层孔隙度预测中的应用
——以准噶尔盆地车排子地区白垩系为例

张宇航[1,2,3],时保宏[1,2],张曰静[4],石好果[4],文雯[5],张杨[5]

1.西安石油大学地球科学与工程学院,西安 710065
2.西安石油大学陕西省油气成藏地质学重点实验室,西安 710065
3.中国石油大学(北京)油气资源与探测国家重点实验室,北京 102249
4.中国石化胜利油田分公司勘探开发研究院,山东东营 257061
5.中国石油青海油田分公司采油一厂,甘肃敦煌 736202

摘 要 准噶尔盆地车排子地区白垩系储层以滩坝相沉积为主,储层砂体薄,纵向变化快,孔隙度估算难度较大。基于Xgboost机器学习算法,根据取心井的岩心实测数据,结合其对应的测井数据,建立了测井孔隙度模型。结果表明,研究区对储层孔隙度影响较大的测井变量为自然伽马测井、声波测井、密度测井和冲洗带电阻率测井,其相关系数分别为0.38、0.42、0.28和0.32。基于特征测井数据,利用Xgboost算法预测的孔隙度与实测孔隙度吻合度较高,相关系数为0.92,均方差为0.20。此外,对近期钻探的新井储层孔隙度进行预测,结果表明孔隙度较高的井段与试油数据相吻合,从侧面反映了模型的可靠性。这一结果为研究区油气藏评价和后期油藏模型的建立提供基础数据,有利于提高研究区勘探的精度。同时,该模型也可用于类似滩坝相、砂体薄的沉积背景下储层孔隙度估算研究。

关键词 机器学习;孔隙度估算;滩坝相;白垩系;车排子凸起

第一作者简介 张宇航,男,1983年出生,博士,讲师,含油气盆地分析,E-mail: yhzhang@xsyu.edu.cn

中图分类号 P613.18 **文献标志码** A

0 引言

岩石孔隙度是指储集岩中孔隙体积与岩石体积的比值[1]。有效孔隙度是指岩石中互相连通孔隙的体积与岩石体积的比值,是油气藏评价和开发阶段普遍关注的物性参数[2]。常规储集岩的孔隙度值一般介于5%~30%,平均为15%[3]。迄今为止,在油气勘探开发阶段中,确定孔隙度的方法主要有两种:一种是直接法,即通过钻井取心在实验室进行系列实验所得,包括渗吸、压汞和气体膨胀等方法[4-5];另一种是间接法,即利用地球物理资料来估算孔隙度[6-7]。这两种方法在实际试油生产和研究中都发挥着不可替代的作用。直接测定法是目前确定孔隙度最准确的方法,为储层研究提供最直接的参考资料。然而,由于钻井的成本较高,通常可供分析的样品较少。另外,实验室分析是相对耗时、昂贵的。间接法主要有常规测井解释方法和人工智能方法。常规的方法主要包括反演法、经验公式法和多元回归方法,这些方法属于线性方法,优点是原理简单、操作简易,但往往误差较大、预测效果不理想,所构建的经验公式在不同油田中泛化能力也较差。

近年来,随着人工智能技术的快速发展,机器学习方法被广泛应用于各个领域[8]。利用机器学习方法,通过建立测井数据与实验室测定数据建立关系,大量学者对储集层物性参数进行预测。侯贤沐等[9]利用不同的机器学习方法对碳酸盐岩储集层的孔隙度和渗透率进行预测,输入参数对预测结果具有明显影响。Nourani *et al.*[10]将机器学习和X射线荧光元

收稿日期:2022-06-08;**修回日期**:2022-08-03;**录用日期**:2022-10-27;**网络出版日期**:2022-10-27

基金项目:国家自然科学基金项目(41711530128);陕西省自然科学基金项目(2021JQ-587);油气资源与探测国家重点实验室开放课题基金(PRP/open-1609)[**Foundation**: National Natural Science Foundation of China, No. 41711530128; Natural Science Foundation of Shaanxi Province, No. 2021JQ-587; Open Project Fund of State Key Laboratory of Petroleum Resources and Exploration, No. PRP/open-1609]

素分析相结合，建立了一种快速、可靠的储层孔隙度预测新技术并被广泛应用。Ahmadi *et al.*[11]基于混合机器学习模型，通过测井数据估算储层孔隙度和渗透率，为油藏建模提供了准确数据。单一线性回归方法预测孔隙度效果不理想，但是基于线性回归与神经网络组合的模型，孔隙度预测效果远远高于线性回归效果[12]。考虑到岩性在一定程度上能够反映孔隙度，因此前人通过建立不同岩性储层类型的测井与储层孔隙度之间的关系，预测不同储层的孔隙度，极大提高了预测准确性[13-15]。上述研究实例主要基于单一机器学习方法，预测模型的泛化能力较差[13]。与其他方法相比，Xgboost模型基于梯度提升决策树算法，在非线性映射问题具有较强处理能力，同时具有更高的预测精度和更强的泛化性能。目前，Xgboost算法被用于估算致密砂岩孔隙度[16-17]、渗透率[18-19]和预测岩性剖面[20-21]。上述方法主要被用于常规大套碎屑岩、碳酸盐岩和致密砂岩中，针对厚度薄变化快的薄储层研究相对薄弱。

由于深层钻井成本的增高和浅层勘探效率较高，浅层圈闭成为油气勘探和增产增储的新目标[22]。准噶尔盆地车排子地区白垩系储层埋藏较浅，是该地区重要的勘探层位[23]。白垩系储层沉积环境主要为滩坝相，呈现泥包砂、砂体薄、纵向变化快的特征。这给估算储层孔隙度带来巨大挑战[24-27]。本文拟基于机器学习方法，建立测井数据与实测数据关系，预测储层孔隙度，为后续储层参数建模和开发方案的设计提供可靠数据支撑。

1 区域地质概况

准噶尔盆地位于我国西北部，是富含油气区带之一。车排子凸起位于准噶尔盆地西北缘的南端，是西部隆起的一个二级构造单元，面积约为1.08×10^4 km^2。车排子凸起与四棵树凹陷、昌吉凹陷、中拐凸起、扎伊尔山相邻，大致呈倒三角形[28]。研究区深部断裂断距大、浅层断距小，为车排子凸起区的主断裂体系，同时发育的EW向正断层[29-31]（图1a）。

随着勘探的不断深入，根据地震资料分析，凸起主体部位在石炭系基底上（图1b），自下而上发育了白垩系下统吐谷鲁群、古近系、新近系沙湾组、塔西河组、独山子组及第四系（图1c）。但是，由于剥蚀严重，造成二叠系、三叠系、侏罗系缺失[29]（图1c）。前人研究表明，车排子凸起发现的原油主要是由昌吉凹陷和四棵树凹陷两大生烃中心的烃源岩，具有双向供源，油源条件比较充足[32-34]。研究区在石炭系火山

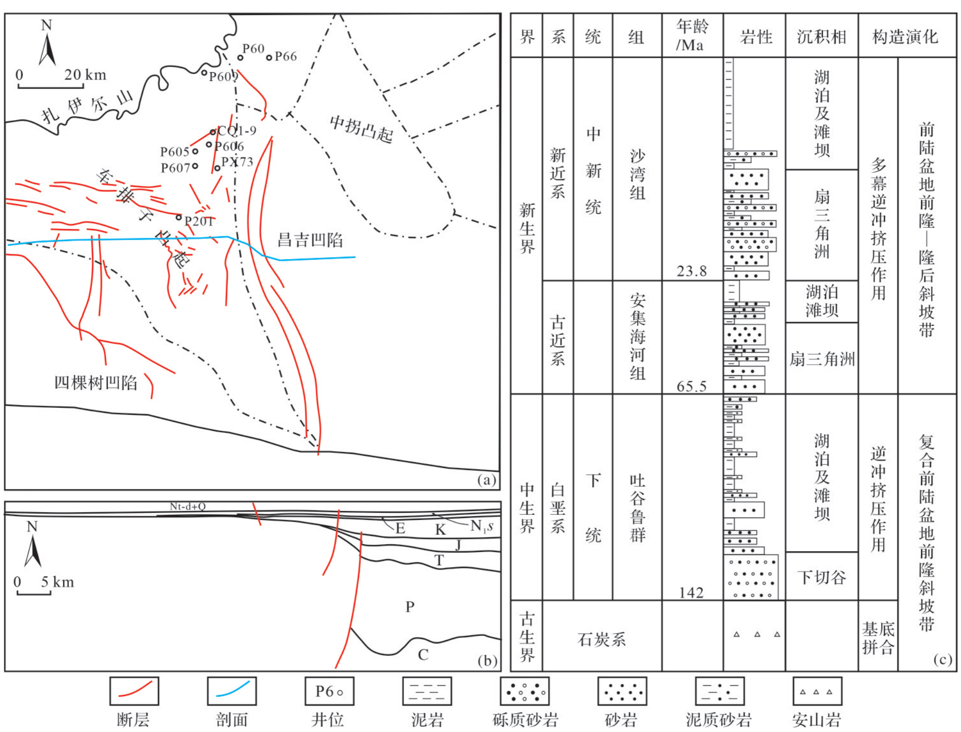

图1 研究区（a）区域地质、（b）构造剖面及（c）综合柱状图

Fig.1 (a) Regional geological characteristics; (b) representative geological profiles; and (c) composite geological histogram of the study area

岩储层获得重大突破[35-36]。浅层主力油层为侏罗系八道湾组、白垩系吐谷鲁群和新近系沙湾组,与深部烃源岩形成下生上储型油气藏[37]。

通过对研究区层序地层学分析,白垩系主要被划分为4个三级层序。每个层序内部岩性主要发育滨浅湖泥岩和滩坝砂,其中泥岩厚度较大,砂岩厚度较小,为2~10 m,泥地比较高。研究区仅最底部层序发育部分扇三角洲沉积,后期湖平面先升高后降低,但总体岩性还是以泥岩为主,仅发育薄砂体[26-27]。因此,研究对象白垩系主要储层为滩坝砂体,具有厚度薄、纵向变化快、横向连续性非常差的特点。

2　实验与方法

2.1　数据来源

车排子地区白垩系中广泛分布的相对薄的滩坝相砂岩是非常重要且有前景的油气勘探层系[38]。然而,储层规模、质量和分布特征深入认识的缺乏,严重限制了勘探突破。估算储层孔隙率是提高砂岩储层有利区域预测的基础。基于此,收集62个白垩系砂岩样品在实验室进行孔隙率测定,以及取样井的生产数据,与估计的孔隙率进行比较。此外,车排子地区7口井(井位如图1a)的常规测井被用来估算白垩系砂岩储层的岩石物理特性(表1)。搜集的测井数据的纵向分辨率为0.125 m左右。通过对比取心岩性和测井特征之间的关系,将取心井段深度回归到相应测井曲线深度。文中数据表所列出的均为回归后的测井深度数据。根据GB/T 29172—2012方法[38],使用QK-98气相孔隙度测量仪测量岩心的孔隙率。通过清洗岩心样品并干燥后,使用波义尔定律测量所有岩心样品的孔隙度。

2.2　Xgboost算法

集成学习算法将多个机器学习算法整合起来,提高预测精度和效果。Xgboost算法就属于集成学习算法的一种。由于其在处理非线性回归问题的优异表现,Xgboost算法被用于各个领域。Chen *et al*.[39]首次提出Xgboost模型。Xgboost算法的主要思想是将集合多个简单的树模型,然后迭代生成一棵新树,来提高模型的精度。它被看作是泛函的最优化,其目标函数为:

$$L(\Phi)=\Sigma_i l\left(\hat{y}_i, y_i\right)+\Sigma_k \Omega\left(f_k\right) \tag{1}$$

表1　研究区部分实测孔隙度及其对应测井数据表

Table 1　Measured porosity and corresponding logging data in the study area

井号	数量/块	深度/m	GR/API	SP/mV	CAL/m	AC/(μs/m)	CNL/%	DEN/(g/cm^3)	RXO/Ω	RT/Ω	实测孔隙度/%	预测孔隙度/%
CQ1-9	7	890.53	67.89	70.27	9.83	126.06	41.26	2.15	1.41	0.96	37.60	38.79
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		912.38	61.76	62.19	9.85	124.98	45.38	2.11	1.99	1.35	37.20	37.20
P60	4	579.70	50.44	128.13	8.99	73.10	59.19	2.06	3.52	3.61	34.80	33.43
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		581.10	50.44	128.13	8.99	72.44	71.28	1.99	4.11	4.17	27.70	33.43
P66	8	299.45	68.21	74.57	13.55	143.20	41.67	2.13	1.56	1.78	35.90	35.90
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		950.40	29.42	69.95	9.23	69.38	27.56	2.40	5.59	5.99	15.00	15.00
P605	7	389.60	233.97	89.96	9.58	130.56	16.77	2.45	2.35	2.37	7.10	7.10
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		391.55	162.42	77.39	10.29	165.08	33.29	2.14	1.86	1.84	6.90	6.90
P606	4	961.00	90.68	100.49	11.12	124.76	40.49	2.14	1.75	1.74	26.90	26.90
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		962.65	76.55	95.36	10.25	112.44	30.51	2.23	7.37	3.98	7.20	7.20
P607	3	344.20	67.15	78.67	10.52	153.06	50.82	2.09	2.23	1.43	40.80	36.93
		365.50	68.51	79.48	11.91	116.58	44.63	2.23	5.05	3.36	40.10	34.37
		366.60	76.05	81.74	11.30	131.02	43.98	2.12	3.37	4.30	40.20	40.20
P609	29	226.19	182.48	33.90	10.31	125.36	17.86	2.31	36.37	1.91	3.30	3.30
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		317.95	72.62	44.24	10.58	122.42	38.47	2.29	4.40	1.18	8.30	8.30

注:GR. 自然伽马;SP. 自然电位;CAL. 井径;AC. 声波时差;CNL. 中子;DEN. 密度;RXO. 冲洗带电阻率;RT. 地层电阻率。

式中：$l\left(\hat{y}_i,y_i\right)$为误差函数(loss函数)，$\Omega\left(f_k\right)$为正则化部分。

误差函数中的$\hat{y}_i$为模型的输出项。正则化部分主要被用来降低模型的复杂度，防止过拟合，提高模型的泛化能力。其函数表达式具体为：

$$\Omega\left(f\right)=\gamma T+\frac{1}{2}\lambda\ /\!/\ \omega\ /\!/\ ^2 \tag{2}$$

式中：T为树的节点个数，ω为节点数。

XGBoost算法与传统的提升算法不同的是，其损失函数通过二阶泰勒展开，即loss函数的二阶展开：

$$\begin{aligned} L^{(t)} &= \Sigma_i l\left(y_i,\hat{y}_i^{(t)}\right)+\Sigma_{i=1}^{t}\Omega\left(f_i\right) \\ &= \Sigma_i l\left(y_i,\hat{y}_i^{(t-1)}+f_t\left(x_i\right)\right)+\Omega\left(f_t\right) \\ &= \Sigma_i\left[g_i f_t\left(x_i\right)+\frac{1}{2}h_i f_t^2\left(x_i\right)\right]+\Omega\left(f_t\right) \end{aligned} \tag{3}$$

式中：g_i为一阶导数，h_i为二阶导数。其函数表达式分别为：

$$g_i=\partial_{\hat{y}_{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{4}$$

$$h_i=\partial^2_{\hat{y}_{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{5}$$

如果定义第j个节点的结合为I_j，令$\mathrm{G}_j=\Sigma_{i\epsilon I_j}g_i$，$\mathrm{H}_j=\Sigma_{i\epsilon I_j}h_i$，并带入(2)式和(3)式，简化得到：

$$L^{(t)}=\Sigma_{j=1}^{T}\left[\mathrm{G}_j\omega_j+\frac{1}{2}\left(\mathrm{H}_j+\lambda\right)\omega^2{}_j\right]+\gamma T \tag{6}$$

求取上式导数，则$\omega^*{}_j=-\dfrac{\mathrm{G}_j}{\mathrm{H}_j+\lambda}$，然后带入(6)式。得：

$$\tilde{L}^{(t)}=-\frac{1}{2}\Sigma_{j=1}^{T}\left(\frac{\mathrm{G}^2{}_j}{\mathrm{H}_j+\lambda}\right)+\gamma T \tag{7}$$

3 结果与讨论

3.1 相关性分析

利用特征工程，优选对孔隙度影响较大的测井变量。从常规测井参数与实测孔隙度相关系数热力图(图2a)可以看出，岩性测井中，自然伽马(GR)测井与孔隙度相关性较好，相关系数为0.38，而自然电位(SP)与井径测井(CAL)对孔隙度影响较小，这两种测井曲线可以很好地反映岩性，对孔隙度不敏感。常规孔隙度测井系列主要包括声波测井(AC)、中子孔隙度测井(AC)和密度测井(DEN)。AC和CNL(补偿中子测井)对孔隙度有显著影响，相关系数分别为0.42和0.28(图2a)，而DEN与孔隙度的相关性较小。电阻率测井系列冲洗带电阻率(RXO)对孔隙度较敏感(图2a)。

根据测井变量与孔隙度之间的相关关系，计算测井对孔隙度的贡献率(图2b)。从图中可以看出，GR、AC、CNL和RXO四种测井对孔隙度的贡献率在0.2左右，说明四种测井对孔隙度的影响较大。因此，将GR、AC、CNL和RXO四条测井曲线用来预测孔隙度。

3.2 预测效果

图3显示了实验室测定的岩心孔隙度值和Xgboost算法预测的孔隙度值的交汇图。为了检验所建立的孔隙度模型的应用情况，并对其准确性进行比较，计算了相关系数。在交会图中，越靠近单位斜率线的点，与实际数据的偏差越小。从图3可以看出，无论是训练集还是测试集，机器学习预测的数据点大多位于单位斜率线附近，验证了其预测孔隙度的准确性较高。

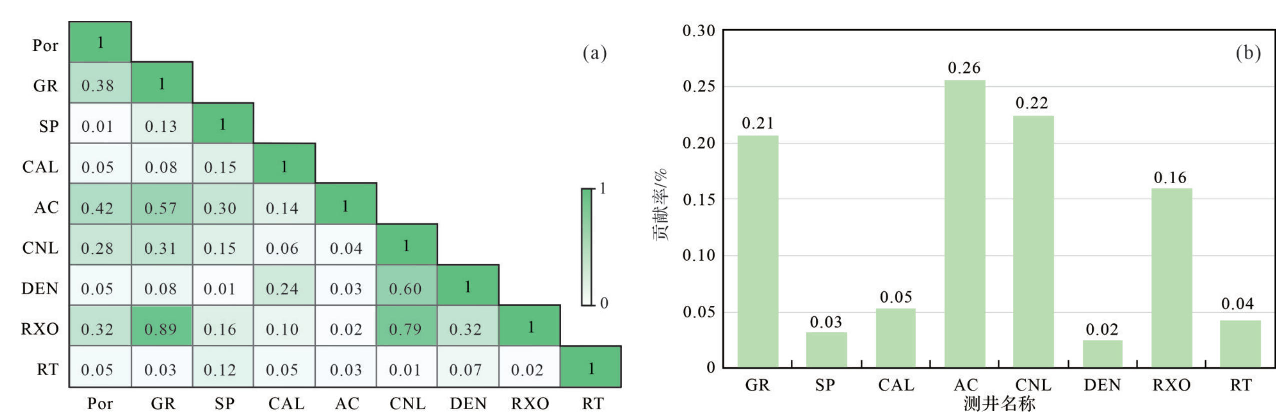

图2 研究区(a)实测孔隙度与测井变量相关系数热力图及(b)测井变量贡献率

Fig.2 (a) Heat map between measured porosity and well log; (b) relative variance contribution for well logs in the study area

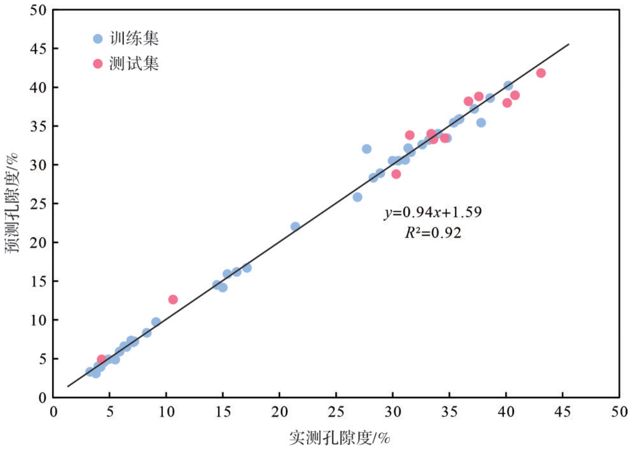

图3　研究区实测与预测孔隙度交会图

Fig.3　Cross plot of measured and predicted porosity in the study area

3.3　单井分析

通过对取心井较多的P609井综合预测,结果表明预测孔隙度与实测孔隙度基本一致(图4)。井段孔隙度预测值与实测值的走势相同,特别是深度在215~225 m之间,实测孔隙度走势先上升再下降,模型预测值反映出这个趋势。与此同时,也有些误差较大的点存在,例如在300 m井段左右,可能是由于孔隙度与测井变量之间的相关性较差,导致该区间预测存在误差值。但是,与常规方法相比,测井解释孔隙度曲线只有孔隙度较高时(如230 m、290 m、310 m左右)与实测孔隙度值较吻合,其他井段存在明显差别,不能反映真实孔隙度;另一方面,测井解释孔隙度的分辨率相对较低,例如在200~210 m和225~235 m井段,测井解释只是解释为一大套储层。总体而言,模型的准确度和可靠性较高,能够提供研究的层位的孔隙度,特别是砂体较薄,纵向变化较快的区域,同样具有较好的预测效果。但模型的精仍有提升空间,可进一步被优化。

为进一步验证模型的准确性,将油田实际试油生产数据作为佐证。2018年,P646井区新增预测储量达1 236.98万吨[37]。2019年,新完钻的PX73井(图1 a)获得商业油流。在PX73井试油井段,1 534.60~1 536.50 m和1 538.50~1 540.00 m两段日产油高达9.01 m³,累计产油约为74.24 m³。根据GR、AC、CNL和RXO测井数据,预测了PX73井孔隙度剖面(图5)。由图可知,在试油成果较高的井段,其对应的孔隙度

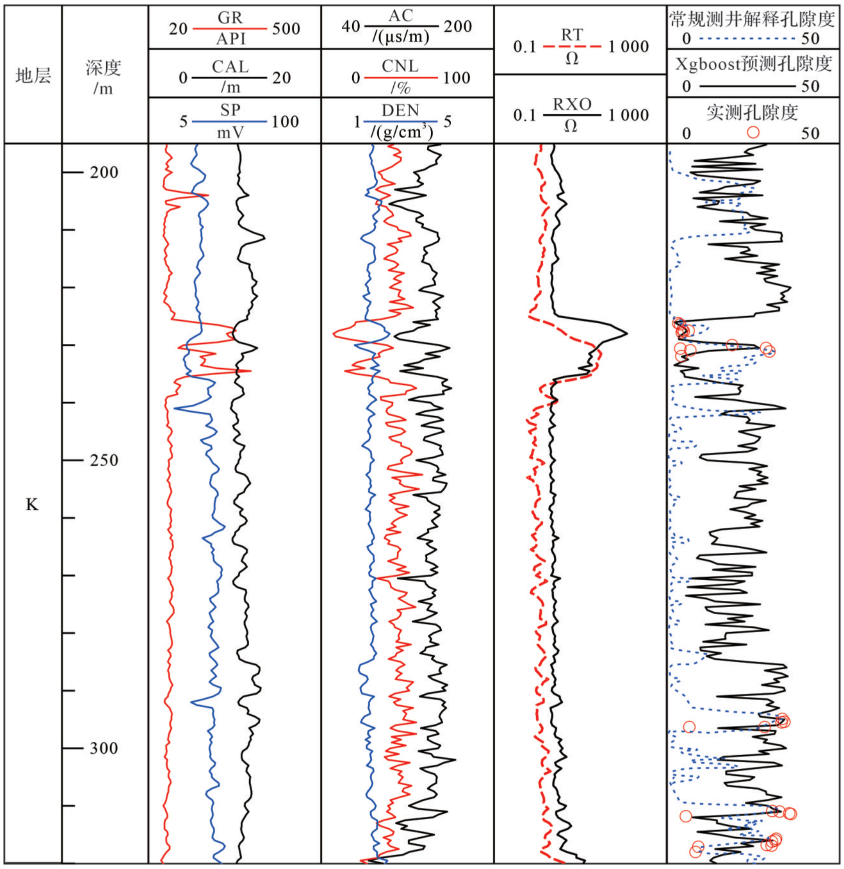

图4　研究区P609井预测孔隙度柱状图

Fig.4　Predicted porosity profile of well P609 in the study area

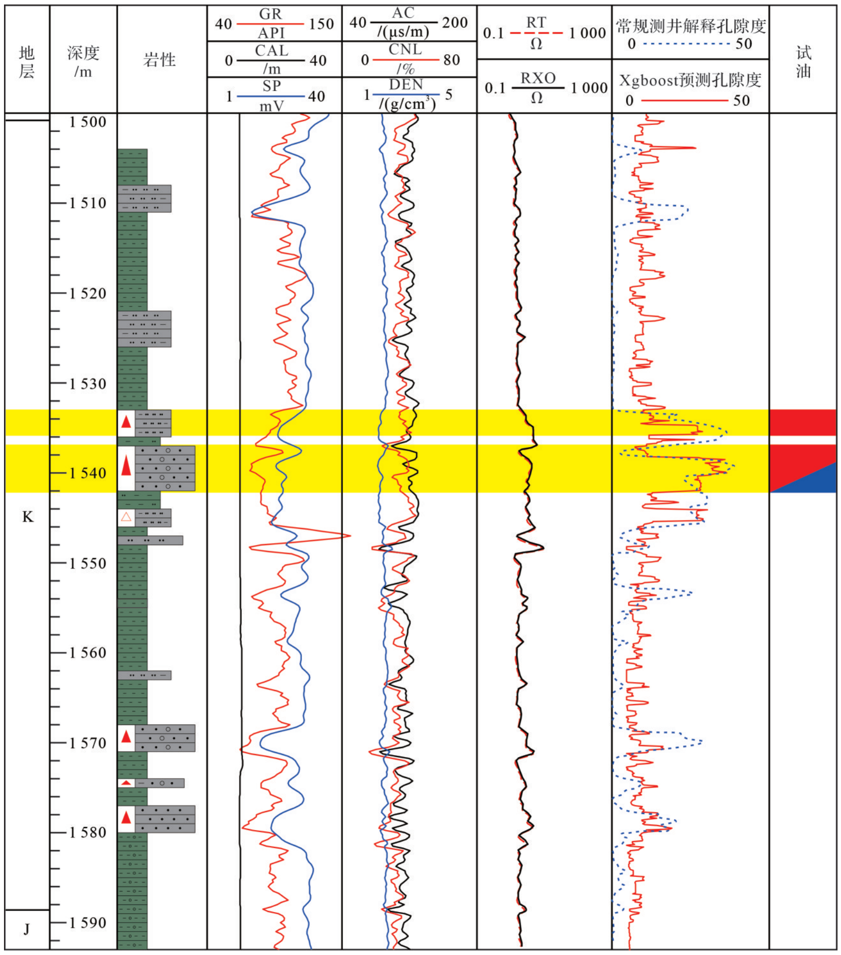

图5 研究区PX73井预测孔隙度柱状图

Fig.5 Predicted porosity profile and production test results of well PX73 in the study area

也较高，说明所建立模型的可靠性较强，同时与测井解释孔隙度曲线进行对比发现，孔隙度总体变化趋势较一致。预测模型能够为油藏评价和油藏模拟提供可靠的数据。

4 结论

（1）机器学习预测在厚度薄、纵向变化快的储层中是可行的。研究表明，以实测孔隙度和测井数据为基础的机器学习方法，在预测非取心剖面储层孔隙度方面优于常规测井解释的方法。

（2）Xgboost算法在处理孔隙度与测井变量之间具有非线性映射关系，表现出较好的预测能力和泛化能力。

（3）本文建立的Xgboost孔隙度预测模型，在薄砂体储层中具有良好的应用效果，可为研究区油气藏评价提供可靠数据，同时也为相似沉积背景下储层孔隙度提供参考依据。

致谢 审稿专家和编辑的有益意见和建议极大地提高了论文的质量，胜利油田勘探开发研究院提供了基础数据和样品，在此一并感谢。

参考文献（References）

[1] Taud H, Martinez-Angeles R, Parrot J F, et al. Porosity estimation method by X-ray computed tomography[J]. Journal of Petroleum Science and Engineering, 2005, 47(3/4): 209-217.

[2] Robin V, Sardini P, Mazurier A, et al. Effective porosity measurements of poorly consolidated materials using non-destructive methods[J]. Engineering Geology, 2016, 205: 24-29.

[3] 贾承造，邹才能，李建忠，等. 中国致密油评价标准、主要类型、基本特征及资源前景[J]. 石油学报，2012，33(3)：343-350. [Jia Chengzao, Zou Caineng, Li Jianzhong, et al. Assessment criteria, main types, basic features and resource prospects of the tight oil in China[J]. Acta Petrolei Sinica, 2012, 33(3): 343-350.]

[4] Fusi N, Martinez-Martinez J. Mercury porosimetry as a tool for improving quality of micro-CT images in low porosity carbonate rocks[J]. Engineering Geology, 2013, 166: 272-282.

［5］ Jian K，Fu X H，Ding Y M，et al. Characteristics of pores and methane adsorption of low-rank coal in China［J］. Journal of Natural Gas Science and Engineering，2015，27：207-218.

［6］ Xiao L，Mao Z Q，Li G R，et al. Calculation of porosity from nuclear magnetic resonance and conventional logs in gas-bearing reservoirs［J］. Acta Geophysica，2012，60(4)：1030-1042.

［7］ Miah M I. Porosity assessment of gas reservoir using wireline log data：A case study of Bokabil Formation，Bangladesh［J］. Procedia Engineering，2014，90：663-668.

［8］ Reichstein M，Camps-Valls G，Stevens B，et al. Deep learning and process understanding for data-driven Earth system science［J］. Nature，2019，566(7743)：195-204.

［9］ 侯贤沐，王付勇，宰芸，等．基于机器学习和测井数据的碳酸盐岩孔隙度与渗透率预测［J］．吉林大学学报（地球科学版），2022，52(2)：644-653.［Hou Xianmu，Wang Fuyong，Zai Yun，et al. Prediction of carbonate porosity and permeability based on machine learning and logging data［J］. Journal of Jilin University (Earth Science Edition)，2022，52(2)：644-653.］

［10］ Nourani M，Alali N，Samadianfard S，et al. Comparison of machine learning techniques for predicting porosity of chalk［J］. Journal of Petroleum Science and Engineering，2022，209：109853.

［11］ Ahmadi M A，Chen Z X. Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs［J］. Petroleum，2019，5(3)：271-284.

［12］ 项云飞，康志宏，郝伟俊，等．基于线性回归与神经网络的储层参数预测复合方法［J］．科学技术与工程，2017，17(31)：46-52.［Xiang Yunfei，Kang Zhihong，Hao Weijun，et al. A composite method of reservoir parameter prediction based on linear regression and neural network［J］. Science Technology and Engineering，2017，17(31)：46-52.］

［13］ 段友祥，王言飞，孙歧峰．选择性集成学习模型在岩性—孔隙度预测中的应用［J］．科学技术与工程，2020，20(3)：1001-1008.［Duan Youxiang，Wang Yanfei，Sun Qifeng. Application of selective ensemble learning model in lithology-porosity prediction［J］. Science Technology and Engineering，2020，20(3)：1001-1008.］

［14］ 邓社根，滕新保，华桂钱，等．砂泥岩储层孔隙度预测的改进模型［J］．黑龙江科技大学学报，2017，27(6)：621-625.［Deng Shegen，Teng Xinbao，Hua Guiqian，et al. A novel prediction model for porosity in sand shale reservoir［J］. Journal of Heilongjiang University of Science & Technology，2017，27(6)：621-625.］

［15］ 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)：450-454.［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)：450-454.］

［16］ 闫星宇，顾汉明，肖逸飞，等．XGBoost算法在致密砂岩气储层测井解释中的应用［J］．石油地球物理勘探，2019，54(2)：447-455.［Yan Xingyu，Gu Hanming，Xiao Yifei，et al. XGBoost algorithm applied in the interpretation of tight-sand gas reservoir on well logging data［J］. Oil Geophysical Prospecting，2019，54(2)：447-455.］

［17］ Pan S W，Zheng Z C，Guo Z，et al. An optimized XGBoost method for predicting reservoir porosity using petrophysical logs［J］. Journal of Petroleum Science and Engineering，2022，208：109520.

［18］ Gu Y F，Zhang D Y，Bao Z D. A new data-driven predictor，PSO-XGBoost，used for permeability of tight sandstone reservoirs：A case study of member of chang 4+5，western Jiyuan oilfield，Ordos Basin［J］. Journal of Petroleum Science and Engineering，2021，199：108350.

［19］ Zhao X B，Chen X J，Huang Q，et al. Logging-data-driven permeability prediction in low-permeable sandstones based on machine learning with pattern visualization：A case study in Wenchang A Sag，Pearl River Mouth Basin［J］. Journal of Petroleum Science and Engineering，2022，214：110517.

［20］ Gu Y F，Zhang D Y，Bao Z D. Lithological classification via an improved extreme gradient boosting：A demonstration of the Chang 4+5 member，Ordos Basin，northern China［J］. Journal of Asian Earth Sciences，2021，215：104798.

［21］ Han R Y，Wang Z W，Wang W H，et al. Lithology identification of igneous rocks based on XGboost and conventional logging curves，a case study of the eastern depression of Liaohe Basin［J］. Journal of Applied Geophysics，2021，195：104480.

［22］ 谢玉洪．低油价背景下中国海油油气勘探进展与发展思考［J］．中国海上油气，2021，33(1)：1-12.［Xie Yuhong. Progress and thinking of CNOOC oil and gas exploration under the background of low oil prices［J］. China Offshore Oil and Gas，2021，33(1)：1-12.］

［23］ 王金铸，王学忠．车排子斜坡带岩性油藏的高效勘探［J］．特种油气藏，2013，20(2)：20-24.［Wang Jinzhu，Wang Xuezhong. Efficient exploration of lithologic reservoirs in the Chepaizi slope zone of western Junggar Basin［J］. Special Oil & Gas Reservoirs，2013，20(2)：20-24.］

［24］ 吕世超．车排子东缘沙湾组薄互层油藏分布特征［J］．科学技术与工程，2020，20(5)：1734-1739.［Lü Shichao. Thin interlayer reservoir distribute character of Shawan Formation in eastern Chepaizi area［J］. Science Technology and Engineering，2020，20(5)：1734-1739.］

［25］ 张曰静，张奎华，隋风贵，等．准噶尔盆地车排子凸起白垩系层序—古地貌耦合控砂机制与砂体预测［J］．东北石油大学学报，2020，44(6)：1-11.［Zhang Yuejing，Zhang Kuihua，Sui Fenggui，et al. Coupling mechanism for sand control and sand body prediction of Cretaceous sequence and paleogeomorphology in the Chepaizi uplift，Junggar Basin［J］. Journal of Northeast Petroleum University，2020，44(6)：1-11.］

［26］ 赵东娜，朱筱敏，董艳蕾，等．地震沉积学在湖盆缓坡滩坝砂

体预测中的应用：以准噶尔盆地车排子地区下白垩统为例[J]. 石油勘探与开发，2014，41（1）：55-61.［Zhao Dongna，Zhu Xiaomin，Dong Yanlei，et al. Application of seismic sedimentology to prediction of beach and bar sandbodies in gentle slope of lacustrine basin：A case study of the Lower Cretaceous in Chepaizi area，Junggar Basin，NW China[J]. Petroleum Exploration and Development，2014，41(1)：55-61.］

［27］ 赵东娜，朱筱敏，董艳蕾，等. 准噶尔盆地车排子地区下白垩统层序地层格架及主控因素分析[J]. 沉积学报，2013，31(6)：1070-1080.［Zhao Dongna，Zhu Xiaomin，Dong Yanlei，et al. Sequence stratigraphic framework of Lower Cretaceous and its main controlling factors in Chepaizi area，Junggar Basin[J]. Acta Sedimentologica Sinica，2013，31(6)：1070-1080.］

［28］ 杨勇，陈世悦，向奎，等. 准噶尔盆地西北缘车排子地区下白垩统层序地层与沉积演化[J]. 中国石油大学学报（自然科学版），2011，35（5）：20-26.［Yang Yong，Chen Shiyue，Xiang Kui，et al. Sequence stratigraphy and sedimentary evolution of Lower Cretaceous series in Chepaizi area，northweastern margin of Junggar Basin[J]. Journal of China University of Petroleum，2011，35(5)：20-26.］

［29］ 董大伟，李理，王晓蕾，等. 准噶尔盆地西缘车排子凸起构造演化及断层形成机制[J]. 吉林大学学报（地球科学版），2015，45（4）：1132-1141.［Dong Dawei，Li Li，Wang Xiaolei，et al. Structural evolution and dislocation mechanism of western margin Chepaizi uplift of Junggar Basin[J]. Journal of Jilin University（Earth Science Edition），2015，45(4)：1132-1141.］

［30］ 胡秋媛，董大伟，赵利，等. 准噶尔盆地车排子凸起构造演化特征及其成因[J]. 石油与天然气地质，2016，37（4）：556-564.［Hu Qiuyuan，Dong Dawei，Zhao Li，et al. Tectonic evolutionary characteristics and their causes of Chepaizi uplift in Junggar Basin[J]. Oil & Gas Geology，2016，37(4)：556-564.］

［31］ 陈石，郭召杰，漆家福，等. 准噶尔盆地西北缘三期走滑构造及其油气意义[J]. 石油与天然气地质，2016，37（3）：322-331.［Chen Shi，Guo Zhaojie，Qi Jiafu，et al. Three-stage strike-slip fault systems at northwestern margin of Junggar Basin and their implications for hydrocarbon exploration[J]. Oil & Gas Geology，2016，37(3)：322-331.］

［32］ 张枝焕，向奎，秦黎明，等. 准噶尔盆地四棵树凹陷烃源岩地球化学特征及其对车排子凸起油气聚集的贡献[J]. 中国地质，2012，39（2）：326-337.［Zhang Zhihuan，Xiang Kui，Qin Liming，et al. Geochemical characteristics of source rocks and their contribution to petroleum accumulation of Chepaizi area in Sikeshu Depression，Junggar Basin[J]. Geology in China，2012，39(2)：326-337.］

［33］ 李二庭，靳军，米巨磊，等. 准噶尔盆地车排子地区原油油源分析[J]. 地球化学，2021，50（5）：492-502.［Li Erting，Jin Jun，Mi Julei，et al. Analysis of crude oil sources in the Chepaizi area，Junggar Basin[J]. Geochimica，2021，50(5)：492-502.］

［34］ 曹剑，胡文瑄，张义杰，等. 准噶尔盆地红山嘴—车排子断裂带含油气流体活动特点地球化学研究[J]. 地质论评，2005，51（5）：591-599.［Cao Jian，Hu Wenxuan，Zhang Yijie，et al. Geochemical analysis on petroleum fluid activity in the Hongshanzui-Chepaizi fault zone，the Junggar Basin[J]. Geological Review，2005，51(5)：591-599.］

［35］ 孟凡超，操应长，崔岩，等. 准噶尔盆地西缘车排子凸起石炭系火山岩储层成因[J]. 中国石油大学学报（自然科学版），2016，40（5）：22-31.［Meng Fanchao，Cao Yingchang，Cui Yan，et al. Genesis of Carboniferous volcanic reservoirs in Chepaizi salient in western margin of Junggar Basin[J]. Journal of China University of Petroleum，2016，40(5)：22-31.］

［36］ 何登发，陈新发，况军，等. 准噶尔盆地石炭系烃源岩分布与含油气系统[J]. 石油勘探与开发，2010，37（4）：397-408.［He Dengfa，Chen Xinfa，Kuang Jun，et al. Distribution of Carboniferous source rocks and petroleum systems in the Junggar Basin[J]. Petroleum Exploration and Development，2010，37(4)：397-408.］

［37］ 宋明水，吕明久，赵乐强，等. 准噶尔盆地车排子凸起油气资源潜力与成藏模式[J]. 中国石油勘探，2016，21（3）：83-91.［Song Mingshui，Lü Mingjiu，Zhao Leqiang，et al. Hydrocarbon potential and accumulation model in Chepaizi uplift，Junggar Basin[J]. China Petroleum Exploration，2016，21(3)：83-91.］

［38］ 中华人民共和国国家质量监督检验检疫总局，中国国家标准化管理委员会. GB/T 29172—2012 岩心分析方法[S]. 北京：中国标准出版社，2013：32-34.［General Administration of Quality Supervision，Inspection and Quarantine of the People's Republic of China，Standardization Administration of the People's Republic of China. GB/T 29172-2012 practices for core analysis[S]. Beijing：Standards Press of China，2013：32-34.］

［39］ Chen T Q，Guestrin C. XGBoost：A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco：ACM，2016：785-794.

Application of Machine Learning for Porosity Estimation of Beach and Bar Sand Bodies in a Lacustrine Basin: A case study of the Lower Cretaceous strata in Chepaizi area, Junggar Basin, NW China

ZHANG YuHang[1,2,3], SHI BaoHong[1,2], ZHANG YueJing[4], SHI HaoGuo[4], WEN Wen[5], ZHANG Yang[5]

1. School of Earth Sciences and Engineering, Xi´an Shiyou University, Xi'an 710065, China

2. Shaanxi Key Lab of Petroleum Accumulation Geology, Xi'an Shiyou University, Xian 710065, China

3. State Key Laboratory of Petroleum Resource and Prospecting, China University of Petroleum(Beijing), Beijing 102249, China

4. Research Institute of Exploration and Development, Shengli Oilfield Company, SINOPEC, Dongying, Shandong 257061, China

5. First Oil Prodction Plant of CNPC Qinghai Oilfield Company, Dunhuang, Gansu 736202, China

Abstract: The facies of Cretaceous reservoirs are beach and bar, and the sandstone reservoirs are characterized by thin and sharp vertical change in the Chepaizi area of the Junggar Basin. As a result, new challenges appear in estimating the porosity of reservoirs. In this study, based on the measured porosity in the laboratory and corresponding logging data, a new porosity estimation model was established using the extreme gradient boosting (Xgboost) machine learning algorithm. The results show that the correlation coefficients between reservoir porosity and GR, AC, CNL, and RXO are 0.38, 0.42, 0.28, and 0.32, respectively, suggesting that porosity is influenced by the logging data in the study area. Based on the input logging data, the predicted porosity using the Xgboost algorithm matches the measured porosity, with a correlation coefficient of 0.92 and a mean squared error of 0.20. To test and verify the predicted results from the Xgboost mothed, we use the production test result as collateral evidence. The results show that the well sections with higher porosity match with the test data, indicating the reliability of the model. This result provides fundamental data for reservoir evaluation and modelling, improving the exploration accuracy in the study area. Furthermore, the model can be used in the study of reservoir porosity estimation in similar sedimentary environments.

Key words: machine learning; porosity estimation; beach and bar facies; Cretaceous; Chepaizi uplift