

文章编号:1009-9603(2022)01-0137-08

DOI:10.13673/j.cnki.cn37-1359/te.2022.01.017

# 基于物理约束数据挖掘算法的 海上油井初期产能预测方法

董银涛<sup>1</sup>,宋来明<sup>1</sup>,张迎春<sup>2</sup>,邱凌<sup>1</sup>,余洋<sup>3</sup>,卢川<sup>1</sup>

(1.中海油研究总院有限责任公司,北京 100028; 2.中联煤层气有限责任公司,  
北京 100026; 3.中国石油勘探开发研究院,北京 100083)

**摘要:**为建立一套海上砂岩油藏定向井初期产能的预测方法,综合考虑地质、工程与开发3个方面17项初期产能的影响因素,基于45口海上砂岩油藏定向井的2700组数据,组合使用Spearman相关系数、随机森林与递归特征消除算法对影响因素进行重要性排序;并结合油藏工程逻辑判别,筛选初期产能的主控因素。选用极端梯度提升(XGBoost)算法构建初期产能预测模型,并基于产能公式改进其损失函数,增强数据挖掘算法的物理约束。结果表明:地层流动系数、孔隙度、层间地层流动系数变异系数、电潜泵下入垂深、储层射开厚度、井眼尺寸、生产压差、电潜泵频率以及油嘴尺寸是海上砂岩油藏定向井初期产能的主控因素。采用物理约束的XGBoost算法对5口井初期产能预测的平均相对误差为9.68%,而无物理约束的XGBoost算法预测初期产能的平均相对误差为11.68%。因此,物理约束可有效提升XGBoost算法对初期产能的预测精度,同时可实现海上砂岩油藏定向井初期产能的准确预测。

**关键词:**定向井;初期产能;主控因素;XGBoost;产能公式

中图分类号:TE319

文献标识码:A

## Initial productivity prediction method for offshore oil wells based on data mining algorithm with physical constraints

DONG Yintao<sup>1</sup>, SONG Laiming<sup>1</sup>, ZHANG Yingchun<sup>2</sup>, QIU Ling<sup>1</sup>, YU Yang<sup>3</sup>, LU Chuan<sup>1</sup>

(1. CNOOC Research Institute Co., Ltd., Beijing City, 100028, China; 2. China United Coalbed Methane Co., Ltd., Beijing City, 100026, China; 3. PetroChina Research Institute of Petroleum Exploration and Development, Beijing City, 100083, China)

**Abstract:** This research aims to establish a prediction method for the initial productivity of directional wells in offshore sandstone oil reservoirs. A total of 17 factors affecting the initial productivity were considered from aspects of the geology, engineering, and development. Given 2700 sets of data from 45 directional wells in offshore sandstone oil reservoirs, the Spearman correlation coefficient, random forest, and recursive feature elimination algorithm were combined to rank the importance of these influencing factors. With the logic of reservoir engineering, the main controlling factors in the initial productivity were selected. On this basis, the initial productivity prediction model was constructed by the extreme gradient boosting (XGBoost) algorithm, and its loss function was improved by referring to the productivity formula to enhance its physical constraints of this data mining algorithm. The results show that the main controlling factors affecting initial productivity of directional wells in offshore sandstone oil reservoirs include the formation flow coefficient, porosity, variation coefficient of the formation flow coefficient between layers, vertical depth of an electric submersible pump (ESP), perforation thickness of reservoirs, borehole size, drawdown, frequency of ESP, and choke size. Furthermore, the average relative error of the XGBoost algorithm with physical constraints for predicting the initial productivity of five wells is 9.68%, and the average relative error of the XGBoost algorithm without physical constraints is 11.68%. Therefore, physical constraints can effectively improve the accuracy of the XGBoost algorithm in productivity prediction, and the proposed method can realize

收稿日期:2021-11-15。

作者简介:董银涛(1991—),男,河南南阳人,工程师,博士,从事油气田开发方面的研究。E-mail:dongyt4@cnooc.com.cn。

基金项目:中海油研究总院有限责任公司博士后基金项目“海上砂岩油藏产能数据挖掘方法研究”(2020-KFZL-015),中海石油(中国)有限公司科技课题“地质油藏大数据深度挖掘及应用研究”(2020-YXKJ-016)。

the accurate prediction of the initial productivity of directional wells in offshore sandstone oil reservoirs.

**Key words:** directional well; initial productivity; main controlling factor; XGBoost; productivity formula

海上砂岩油藏定向井的初期产能预测是油藏开发前期一项基础性和关键性的研究工作<sup>[1]</sup>。油井产能的影响因素众多,包括沉积环境、流固物性等地质因素,完井技术、采油方式等工程因素以及渗流规律、生产制度等开发因素<sup>[2-7]</sup>。为从中选出油井产能的主控因素,使建立的产能预测模型具有更好的可解释性,前人提出了多种产能主控因素的筛选方法。李展峰等借助灰色关联方法,分别确定了海上复杂油藏与致密油藏水平井产能的主控因素<sup>[8-9]</sup>。GAO等利用主成分分析法,指出地质因素与工程因素共同决定着河流相储层的油井产能<sup>[10]</sup>。宋宣毅等联合使用Pearson相关系数与随机森林算法,确定了特低渗透油藏初期产能的主控因素<sup>[11]</sup>。RASTO-GI等综合运用Pearson相关系数、递归特征消除和随机森林等算法,筛选了水力压裂页岩油气井产能的主控因素<sup>[12]</sup>。由此可见,目前研究多通过组合使用多种主控因素筛选算法,并综合其结果以确定产能主控因素,但其筛选过程仅依据数据挖掘算法的计算结果,而缺少基于油藏工程理论的判别与指导。

在基于数据挖掘算法建立产能预测模型的研究中,现已有神经网络<sup>[13-15]</sup>、支持向量回归<sup>[16-17]</sup>、随机森林<sup>[18]</sup>等多种数据挖掘算法得到应用与改进。田亚鹏等采用遗传算法优化神经网络的权值与阈值,提升该算法的产能预测精度<sup>[15,19]</sup>。SONG等使用粒子群算法优化长短时记忆神经网络,对火山岩储层的油井产能进行预测<sup>[20]</sup>。马文礼等混合使用遗传算法与粒子群算法,对支持向量回归的超参数寻优,实现页岩气井产能的预测<sup>[16]</sup>。可见,此类研究的重点在于应用数据挖掘算法预测产能并提升其精度,而如何将数据挖掘算法与油藏工程理论交叉融合,从而增强数据挖掘算法的物理可解释性仍有待深入探索<sup>[21-22]</sup>。为此,HASSAN等基于神经网络确定产能方程的经验系数,建立了数据挖掘算法与油藏工程理论的关联<sup>[22]</sup>。张亦知等以物理知识构建目标函数,从而约束湍流计算公式并提升其精度<sup>[23]</sup>。然而,在初期产能预测领域尚缺乏有效融合数据挖掘算法与油藏工程理论的研究,即数据挖掘算法在初期产能预测领域仍然缺乏物理可解释性。

综上所述,笔者针对海上砂岩油藏定向井的初期产能,通过组合使用Spearman相关系数、随机森林与递归特征消除算法,为其影响因素做重要性排

序,并结合油藏工程理论制定主控因素的筛选标准,从而确定初期产能主控因素。进而基于XGBoost算法,将主控因素作为初期产能预测模型的输入参数,并依据产能公式改进XGBoost算法的损失函数,以增强其物理可解释性,从而实现海上砂岩油藏定向井初期产能预测模型的建立与应用。

## 1 影响因素与数据选取

综合考虑地质、工程与开发3个方面,选取了17项油井初期产能的影响因素。其中静态地质因素包括储层中深、地层流动系数、孔隙度、含油饱和度、层间地层流动系数变异系数和层间地层流动系数极差6项;工程因素包括井眼尺寸、完井方式、储层射开厚度和电潜泵下入垂深4项;动态开发因素包括含水率、气油比、生产压差、井口压力、油嘴尺寸、电潜泵频率和电潜泵入口压力7项。

对于开发因素,综合考虑初期产能的计算标准<sup>[24]</sup>以及开发因素的动态变化对初期产能的影响,选用油井投产后正常生产的前60组日度数据。对于地质因素,通过对各井的油组地质数据按其射开厚度进行加权求和,整理得各项地质因素的单井数值。对于工程因素,采用独热编码方法处理其中的非数值变量,整理得各项工程因素的单井数值。考虑到开发因素与地质、工程因素的数据量级不统一,对各井的地质、工程因素按单井60组的数据量进行复制扩充,与该井开发因素的数据量进行匹配。最终,整理得45口海上砂岩油藏定向井共计2700组相关数据,作为初期产能主控因素筛选与预测模型建立的基础(表1)。

## 2 主控因素筛选

为从17项影响因素中筛选出对初期产能有较大影响的主控因素,将其作为初期产能预测模型的输入参数,本文组合使用Spearman相关系数、随机森林与递归特征消除算法分别评价地质、工程与开发3个方面的初期产能影响因素,并借助油藏工程理论辅助筛选初期产能主控因素。

### 2.1 重要性评价原理

相关系数是反映两个变量间关系密切程度的统计指标,其取值为 $[-1, 1]$ 。相关系数为正则表示

表1 基础数据分布范围  
Table1 Range of basic data distribution

因素	分布范围	因素	分布范围
初期产能/ ( $\text{m}^3 \cdot \text{d}^{-1}$ )	42~805	地层流动系数/ ( $\text{mD} \cdot \text{m} \cdot (\text{mPa} \cdot \text{s})^{-1}$ )	113~2 464
储层中 深/m	1 029~1 347	层间地层流 动系数变异系数	0.76~2.03
孔隙度/%	24~30	层间地层流 动系数极差	26~3 325
含油饱 和度/%	59~80	储层射 开厚度/m	162~802
含水率/%	0~17.5	气油比	0~424
油嘴尺 寸/mm	7~59	井口压 力/MPa	1.1~8.32
生产压 差/MPa	0.97~8.75	电潜泵入 口压力/MPa	2.19~9.81
电潜泵 频率/Hz	30~64	电潜泵下 入垂深/m	798~1 585
井眼 尺寸/m	0.17~0.32	完井方式	套管射孔/压裂 充填/裸眼筛管/ 裸眼砾石充填

两个变量呈正相关,为负则呈负相关,为0则表示两个变量不相关。两个变量的相关系数越趋近于1或-1,则表示其相关关系越强,而越趋近于0表示相关关系越弱。依据数据类型,选用Spearman相关系数并根据计算结果的绝对值评价各因素的重要性,其表达式为:

$$\rho = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \quad (1)$$

随机森林属于集成学习算法,其通过随机采样与随机特征选取构建多棵决策树,并以各决策树回归结果的平均值作为模型的预测结果。该算法通过统计每个特征使不纯度指标降低的程度来完成特征重要性的排序,不纯度降低值越大,则表明该特征越重要<sup>[11,25]</sup>。采用均方误差作为不纯度评价指标,计算各个特征的重要性权重,其表达式为:

$$\beta_i = \frac{\delta_i}{\sum_i \delta_i} \quad (2)$$

递归特征消除算法通过使用一个基模型进行多轮训练,并在每轮训练后将基模型计算得到的特征重要性权重进行排序,移除若干权重低的特征;再以新的特征集进行训练,直至筛选出最重要特征,从而实现特征重要性的排序<sup>[12]</sup>。根据数据类型,选择随机森林算法作为递归特征消除算法的基模型,对各因素进行重要性评价。

## 2.2 主控因素筛选标准

通过上述算法可从数值上确定各项影响因素与初期产能间关系的密切程度并完成重要性评价,但数据挖掘算法无法识别各项影响因素与初期产能间的逻辑关系,所以不能直接将重要性排序结果作为初期产能主控因素筛选的依据。因此,需进一步借助油藏工程理论,从逻辑角度进行判别,即识别各项影响因素与初期产能间的关联是表示其控制着初期产能还是受初期产能控制,从而实现初期产能主控因素的筛选。采用两项逻辑判别原则作为初期产能主控因素筛选的依据:①该因素可影响初期产能,而不受初期产能影响。②该因素在油井投产前应确定其数值。至少满足以上一条原则,则认为该因素可作为初期产能主控因素。

基于以上两项逻辑判别原则,结合初期产能影响因素的重要性排名,分别从地质、工程与开发3个方面中选取重要性综合排名为前3的因素作为初期产能主控因素,以此作为主控因素的筛选标准。

## 2.3 初期产能主控因素

分别采用Spearman相关系数、随机森林与递归特征消除算法评价地质、工程与开发3类影响因素对初期产能影响的重要性,并综合其结果对3类影响因素进行重要性综合排名。由Spearman相关系数与随机森林算法可得重要性权重(图1),据此得出重要性排名,而递归特征消除算法可直接获得重要性排名(表2),对3种算法所得的重要性排名取平均值即为重要性综合排名。由表2可知,地质因素的重要性综合排名为:地层流动系数>孔隙度>层间地层流动系数变异系数>储层中深>含油饱和度>层间地层流动系数极差;工程因素的重要性综合排名为:电潜泵下入垂深>储层射开厚度>井眼尺寸>完井方式;开发因素的重要性综合排名为:生产压差>电潜泵入口压力>电潜泵频率>气油比>油嘴尺寸>井口压力>含水率。

依据主控因素筛选标准(表2),可确定初期产能主控因素分别为地层流动系数、孔隙度、层间地层流动系数变异系数、电潜泵下入垂深、储层射开厚度、井眼尺寸、生产压差、电潜泵频率以及油嘴尺寸。以上9项主控因素即为初期产能预测模型的输入参数,用于预测海上砂岩油藏定向井的初期产能。

## 3 初期产能预测模型

### 3.1 XGBoost算法基本原理

XGBoost算法属于集成算法,通过多个弱学习

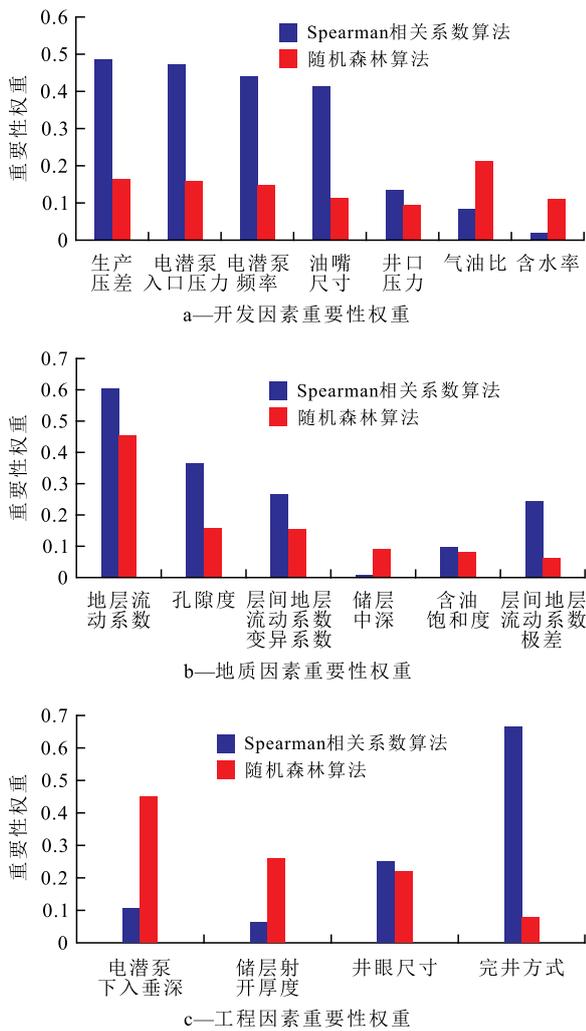


图1 不同算法各因素重要性权重

Fig.1 Importance weights of each factor for different algorithms

器组成一个强学习器,从而实现更精确的预测。该算法的基本原理是以回归树为基础,通过不断添加新的树构建集成回归树模型;每次添加新的树即学习一个新函数,用来拟合上一棵树的计算结果的残差;累加各棵树的计算结果,逐步降低模型整体的残差使其预测值逐渐逼近真实值,实现其回归与预测功能<sup>[26-28]</sup>。累加各棵树的表达式为:

$$\hat{y}_i = \sum_k f_k(x_i) \quad f_k \in F \quad (3)$$

在模型建立过程中,确定各棵树的结构,应以其预测值与真实值的差距最小、模型的复杂度最小为目标。因此,以损失函数与正则项组成该模型的目标函数,其表达式为:

$$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (4)$$

其中:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (5)$$

在模型训练过程中,因采用累加迭代的方式构

表2 初期产能影响因素重要性排名及其主控因素筛选  
Table2 Importance ranking of influencing factors and selection of controlling factors in initial productivity

因素	不同算法重要性排名			重要性综合排名	是否满足原则①	是否满足原则②	是否可作为初期产能主控因素
	Spearman相关系数	随机森林	递归特征消除				
地层流动系数	1	1	1	1	是	是	是
孔隙度	2	2	3	2	是	是	是
层间地层流动系数变异系数	3	3	2	3	是	是	是
储层中深	6	4	4	4	是	是	是
含油饱和度	5	5	5	5	是	是	是
层间地层流动系数极差	4	6	6	6	是	是	是
电潜泵下入垂深	3	1	1	1	是	是	是
储层射开厚度	4	2	2	2	是	是	是
井眼尺寸	2	3	3	2	是	是	是
完井方式	1	4	4	3	是	是	是
生产压差	1	2	1	1	否	是	是
电潜泵入口压力	2	3	2	2	否	否	否
电潜泵频率	3	3	4	3	是	是	是
气油比	6	1	3	3	否	否	否
油嘴尺寸	4	4	6	4	是	是	是
井口压力	5	7	5	5	否	否	否
含水率	7	6	7	6	否	否	否

建集成了回归树模型,故(4)式应使用其迭代形式,同时利用泰勒展开对其损失函数做近似处理,表达式为:

$$L^{(i)} = \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (6)$$

对(6)式求关于 $\omega_j$ 的偏导,可得目标函数在 $\omega_j =$

$-\sum_{i \in I_j} g_i / \left( \sum_{i \in I_j} h_i + \lambda \right)$ 时取得极值,将该值代入(6)式

可得目标函数最优值,用于确定回归树的最佳结构,其表达式为:

$$L^{(i)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (7)$$

但实际应用中,由于无法遍历所有回归树的结

构,故采用贪心算法对比每个节点处每个特征在分裂前后目标函数的变化,其表达式为:

$$L_{\text{split}} = \frac{1}{2} \left[ \frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (8)$$

若(8)式的最大值为正,表明模型预测的误差减小,则该节点向下分裂,反之则停止分裂,如此重复即可完成回归树模型的构建。

### 3.2 损失函数改进

在预测初期产能构建回归树的过程中,是以损失函数最小化为目标即初期产能的预测值与真实值相差最小。定向井产能的表达式为:

$$q = \frac{0.54287KH\Delta pC}{\mu_o B_o \ln\left(\frac{r_e}{r_w} + S\right)} \quad (9)$$

根据(9)式中产能与地层流动系数(KH/μ<sub>o</sub>)及生产压差(Δp)的关系,改进XGBoost算法中的损失函数,为回归树的构建过程增加物理约束,从而增强模型整体的物理可解释性。根据定向井产能公式,令:

$$D = \frac{0.54287C}{B_o \ln\left(\frac{r_e}{r_w} + S\right)} \quad (10)$$

将(10)式代入(9)式可得:

$$q = \frac{KH\Delta p}{\mu_o} D \quad (11)$$

其中,地层流动系数与生产压差在初期产能预测过程中均为输入参数。

根据(11)式,初期产能的预测值与真实值的表达式分别为:

$$\hat{q}_i = \frac{K_i H_i \Delta p_i}{\mu_{oi}} \hat{D}_i \quad (12)$$

$$q_i = \frac{K_i H_i \Delta p_i}{\mu_{oi}} D_i \quad (13)$$

通过令  $\hat{D}_i$  趋近于  $D_i$ , 可得  $\frac{\hat{q}_i}{D_i} - \frac{K_i H_i \Delta p_i}{\mu_{oi}}$  趋近于

0,即可使得预测值趋近于真实值,并达到损失函数最小化的目标,故将损失函数写为:

$$l = \left( \frac{\hat{q}_i}{D_i} - \frac{K_i H_i \Delta p_i}{\mu_{oi}} \right)^2 \quad (14)$$

整理(14)式可得:

$$l = \left( \frac{K_i H_i \Delta p_i}{\mu_{oi}} \right)^2 \left( \frac{\hat{q}_i}{q_i} - 1 \right)^2 \quad (15)$$

对(15)式取一阶导数与二阶导数并代入(7)式,可得依据产能公式构建的目标函数最优值,其表达式为:

$$L^{(o)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left[ \sum_{i \in I_j} 2 \left( \frac{K_i H_i \Delta p_i}{\mu_{oi}} \right)^2 \left( \frac{\hat{q}_i}{q_i} - 1 \right) \frac{1}{q_i} \right]^2}{\sum_{i \in I_j} 2 \left( \frac{K_i H_i \Delta p_i}{\mu_{oi}} \right)^2 \left( \frac{1}{q_i} \right)^2 + \lambda} + \gamma T \quad (16)$$

运用(16)式即可确定回归树的最优结构,完成初期产能预测模型的建立。初期产能预测时,输入各项主控因素,根据(3)式即可得到模型预测的初期产能。

由(15)式与(16)式可知,本文改进的损失函数使XGBoost算法在构建回归树的过程中,既使初期产能预测值趋近于真实值,也使地层流动系数与生产压差保持其在产能公式中的比例关系,具有物理可解释性。

### 3.3 初期产能预测模型构建

基于以上原理与公式,构建初期产能预测模型(图2),具体步骤包括:①将9项主控因素作为模型输入参数,初期产能作为模型预测的目标值,由以上参数组成总数据集,并按8:1的比例划分为训练集与测试集。②基于训练集数据,采用改进的XGBoost算法训练模型,并通过网格搜索方法确定模型的最优超参数,完成初期产能预测模型的构建。③使用测试集数据检验模型预测初期产能的准确性。

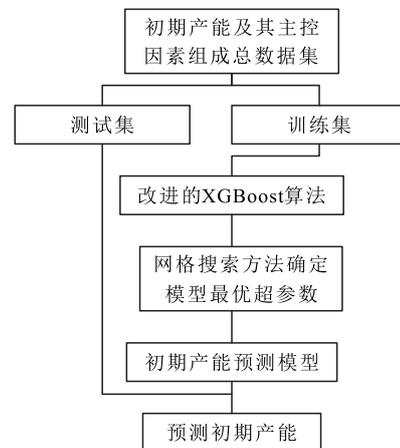


图2 基于改进XGBoost算法的初期产能预测模型构建流程

Fig.2 Construction process of initial productivity prediction model based on improved XGBoost algorithm

### 4 实例应用

依据初期产能预测模型构建流程(图2),利用45口井的初期产能及其9项主控因素作为总数据集,取40口井的2400组数据作为训练集用于模型训练,其余5口井的300组数据作为测试集进行模型检验。

利用训练集数据,通过网格搜索方法确定了基于改进XGBoost算法的初期产能预测模型的最优超参数,由图3可优选出回归树个数为80,树的最大深度为6;再利用上述超参数的值进行训练,建立初期产能预测模型。利用测试集数据对初期产能预测模型进行检验,通过对单井初期的60组日产油量同时进行预测,并对比其预测值与真实值(图4),评价模型的预测精度。因初期产能为投产初期日产油量的平均值,故对单井预测值取平均来评价模型对

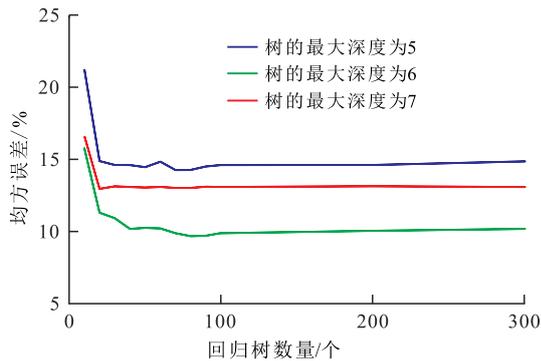


图3 基于改进XGBoost算法的最优超参数网格搜索结果  
Fig.3 Optimal results of hyper-parameter grid search based on improved XGBoost algorithm

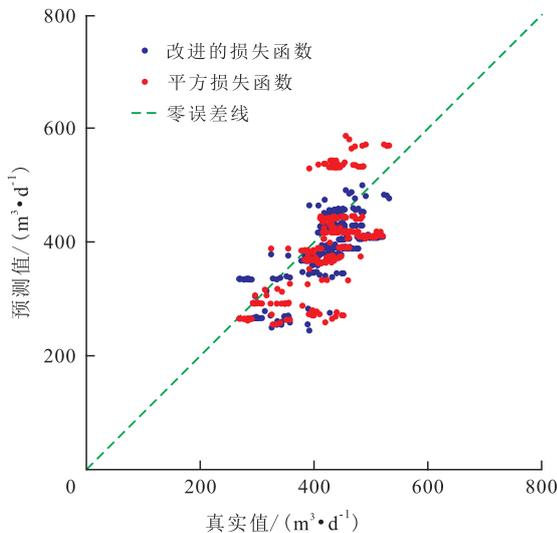


图4 不同损失函数的XGBoost算法初期产能预测模型对比

Fig.4 Comparison of initial productivity prediction models by XGBoost algorithm with different loss functions

其初期产能的预测精度。

为进一步验证基于改进XGBoost算法的初期产能预测模型的准确性,采用相同数据与参数,选择常用的平方损失函数再次建立基于XGBoost算法的初期产能预测模型,并对比2种模型的预测结果(图4)。在采用产能公式改进的损失函数与采用平方损失函数的条件下,对初期产能预测的平均相对误差分别为9.68%和11.68%。可见,基于筛选出的9项初期产能主控因素,并采用XGBoost算法,可准确预测海上砂岩油藏定向井的初期产能;且采用产能公式改进的损失函数增强了算法物理可解释性,同时提升了该算法对初期产能的预测精度。

### 5 结论

综合 Spearman 相关系数、随机森林与递归特征消除算法的结果,分别确定了地质、工程与开发3类影响因素对初期产能影响的重要性排序。基于油藏工程理论建立的初期产能主控因素筛选标准,为数据挖掘算法的重要性评价结果提供了主控因素筛选的逻辑判别依据。据此筛选出的海上砂岩油藏定向井初期产能主控因素为:地层流动系数、孔隙度、层间地层流动系数变异系数、电潜泵下入垂深、储层射开厚度、井眼尺寸、生产压差、电潜泵频率以及油嘴尺寸。物理约束的XGBoost算法预测初期产能的平均相对误差为9.68%,低于无物理约束的XGBoost算法的误差。因此,基于产能公式改进的损失函数,增强了XGBoost算法的物理可解释性,同时提高了其初期产能的预测精度,并实现了海上砂岩油藏定向井初期产能的准确预测。

#### 符号解释

- $B_0$ ——地层原油体积系数,无量纲;
- $C$ ——校正系数,无量纲;
- $d$ ——秩次的差值;
- $d_i$ ——第*i*组数据秩次的差值;
- $D$ ——校正系数,无量纲;
- $D_i$ ——第*i*个节点处校正系数的真实值,无量纲;
- $\hat{D}_i$ ——第*i*个节点处校正系数的预测值,无量纲;
- $f$ ——回归树的输出值;
- $f_k$ ——第*k*棵回归树的输出值;
- $F$ ——回归树的空间;
- $g$ ——损失函数的一阶导数;
- $g_i$ ——第*i*个节点处损失函数的一阶导数;
- $h$ ——损失函数的二阶导数;
- $h_i$ ——第*i*个节点处损失函数的二阶导数;

$H$ ——储层有效厚度, m;  
 $H_i$ ——第  $i$  个节点处储层有效厚度, m;  
 $i$ ——数据序号;  
 $I$ ——总集合;  
 $I_j$ ——第  $j$  个节点处叶节点总集合;  
 $I_L$ ——分裂后左侧子树的总集合;  
 $I_R$ ——分裂后右侧子树的总集合;  
 $j$ ——节点序号;  
 $k$ ——回归树序号;  
 $K$ ——渗透率, mD;  
 $K_i$ ——第  $i$  个节点处渗透率, mD;  
 $l$ ——损失函数;  
 $L$ ——目标函数;  
 $L^{(t)}$ ——第  $t$  个时间步的目标函数;  
 $L_{\text{split}}$ ——判别节点分裂的目标函数;  
 $n$ ——因素的数据量;  
 $\Delta p$ ——生产压差, MPa;  
 $\Delta p_i$ ——第  $i$  个节点处生产压差, MPa;  
 $q$ ——初期产能,  $\text{m}^3/\text{d}$ ;  
 $\hat{q}$ ——初期产能预测值,  $\text{m}^3/\text{d}$ ;  
 $q_i$ ——第  $i$  个节点处初期产能真实值,  $\text{m}^3/\text{d}$ ;  
 $\hat{q}_i$ ——第  $i$  个节点处初期产能预测值,  $\text{m}^3/\text{d}$ ;  
 $r_c$ ——供液半径, m;  
 $r_w$ ——井眼半径, m;  
 $S$ ——表皮系数, 无量纲;  
 $t$ ——时间步;  
 $T$ ——叶节点数量;  
 $x$ ——样本;  
 $x_i$ ——第  $i$  棵回归树的样本;  
 $y$ ——真实值;  
 $\hat{y}$ ——预测值;  
 $y_i$ ——第  $i$  棵回归树的真实值;  
 $\hat{y}_i$ ——第  $i$  棵回归树的预测值;  
 $\beta$ ——重要性权重;  
 $\beta_i$ ——第  $i$  个特征的重要性权重;  
 $\gamma$ ——叶节点数量的惩罚系数, 无量纲;  
 $\delta$ ——不纯度降低值;  
 $\delta_i$ ——第  $i$  个特征的不纯度降低值;  
 $\lambda$ ——叶节点权重的惩罚系数, 无量纲;  
 $\mu_o$ ——地面原油黏度,  $\text{mPa}\cdot\text{s}$ ;  
 $\mu_{oi}$ ——第  $i$  个节点处的地面原油黏度,  $\text{mPa}\cdot\text{s}$ ;  
 $\rho$ ——Spearman 相关系数;  
 $\omega$ ——叶节点权重;  
 $\Omega(f_k)$ ——正则项, 是模型复杂度的惩罚项。

### 参考文献

- [1] 马奎前, 刘英宪, 苏彦春. 渤海 S 油田高含水水期加密井产能评价[J]. 断块油气田, 2013, 20(1): 71-73.
- [2] 谭忠健, 胡云, 张国强, 等. 渤中 19-6 构造复杂储层流体评价及产能预测[J]. 石油钻采工艺, 2018, 40(6): 764-774.
- [3] 李波, 罗宪波, 刘英, 等. 海上稠油油田合理单井产能预测新方法[J]. 中国海上油气, 2008, 20(4): 243-245.
- [4] 张利军, 田冀, 朱国金. 海上断块油田定向井初期产能评价方法分析[J]. 石油钻探技术, 2015, 43(1): 111-116.
- [5] 谢琳, 杨玥, 李杰, 等. 昆北油田切 12 井区基岩油藏产能主控因素及高效开发实践[J]. 油气地质与采收率, 2017, 24(6): 114-119.
- [6] 贾晓飞, 雷光伦, 孙召勃, 等. 三维各向异性油藏水平井产能新公式[J]. 油气地质与采收率, 2019, 26(2): 113-119.
- [7] 林旺, 范洪富, 刘立峰, 等. 工程参数对致密油藏压裂水平井产能的影响[J]. 油气地质与采收率, 2017, 24(6): 120-126.
- [8] 李展峰, 张占女, 王树涛, 等. M1-1 油田厚陡窄复杂油藏水平井产能预测方法研究[J]. 石油地质与工程, 2020, 34(2): 71-75.
- [9] 梁涛, 常毓文, 郭晓飞, 等. 巴肯致密油藏单井产能参数影响程度排序[J]. 石油勘探与开发, 2013, 40(3): 357-362.
- [10] GAO Pengyu, JIANG Chong, HUANG Qin, et al. Fluvial facies res-

- ervoir productivity prediction method based on principal component analysis and artificial neural network[J]. *Petroleum*, 2016, 2(1): 49-53.
- [11] 宋宣毅, 刘月田, 马晶, 等. 基于灰狼算法优化的支持向量机产能预测[J]. *岩性油气藏*, 2020, 32(2): 134-140.  
SONG Xuanyi, LIU Yuetian, MA Jing, et al. Productivity forecast based on support vector machine optimized by grey wolf optimizer[J]. *Lithologic Reservoirs*, 2020, 32(2): 134-140.
- [12] RASTOGI A, SHARMA A. Quantifying the impact of fracturing chemicals on production performance using machine learning[C]. SPE 197095-MS, 2019.
- [13] BUKHAMSEEN N, ERTEKIN T. The use of artificial neural networks to quantify the effect of formation damage on well production response[C]. SPE 188094-MS, 2017.
- [14] 谷建伟, 周梅, 李志涛, 等. 基于数据挖掘的长短期记忆网络模型油井产量预测方法[J]. *特种油气藏*, 2019, 26(2): 77-81, 131.  
GU Jianwei, ZHOU Mei, LI Zhitao, et al. Oil well production forecast with long-short term memory network model based on data mining[J]. *Special Oil & Gas Reservoirs*, 2019, 26(2): 77-81, 131.
- [15] 田亚鹏, 鞠斌山. 基于遗传算法改进BP神经网络的页岩气产量递减预测模型[J]. *中国科技论文*, 2016, 11(15): 1 710-1 715.  
TIAN Yapeng, JU Binshan. A model for predicting shale gas production decline based on the BP neural network improved by the genetic algorithm[J]. *China Sciencepaper*, 2016, 11(15): 1 710-1 715.
- [16] 马文礼, 李治平, 孙玉平, 等. 基于机器学习的页岩气产能非确定性预测方法研究[J]. *特种油气藏*, 2019, 26(2): 101-105.  
MA Wenli, LI Zhiping, SUN Yuping, et al. Non-deterministic shale gas productivity forecast based on machine learning[J]. *Special Oil & Gas Reservoirs*, 2019, 26(2): 101-105.
- [17] 王威. 致密油藏水平井体积压裂初期产能预测[J]. *新疆石油地质*, 2016, 37(5): 575-579.  
WANG Wei. Forecast of initial horizontal well productivity in tight reservoirs by volumetric fracturing process[J]. *Xinjiang Petroleum Geology*, 2016, 37(5): 575-579.
- [18] 李菊花, 陈晨, 肖佳林. 基于随机森林算法的页岩气多段压裂井产量预测[J]. *长江大学学报: 自然科学版*, 2020, 17(4): 34-38.  
LI Juhua, CHEN Chen, XIAO Jialin. Yield prediction of shale gas multi-stage fracturing wells based on random forest algorithm[J]. *Journal of Yangtze University: Natural Science Edition*, 2020, 17(4): 34-38.
- [19] 陈娟, 黄浩勇, 刘俊辰, 等. 基于GA-BP神经网络的长宁地区页岩气水平井产能预测技术[J]. *科学技术与工程*, 2020, 20(5): 1 851-1 858.  
CHEN Juan, HUANG Haoyong, LIU Junchen, et al. Production predicting technology of shale gas fracturing horizontal well in Changning area based on the GA-BP neural network model[J]. *Science Technology and Engineering*, 2020, 20(5): 1 851-1 858.
- [20] SONG Xuanyi, LIU Yuetian, XUE Liang, et al. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model[J]. *Journal of Petroleum Science and Engineering*, 2020, 186: 106682.
- [21] 张凯, 赵兴刚, 张黎明, 等. 智能油田开发中的大数据及智能优化理论和方法研究现状及展望[J]. *中国石油大学学报: 自然科学版*, 2020, 44(4): 28-38.  
ZHANG Kai, ZHAO Xinggang, ZHANG Liming, et al. Current status and prospect for the research and application of big data and intelligent optimization methods in oilfield development[J]. *Journal of China University of Petroleum: Edition of Natural Science*, 2020, 44(4): 28-38.
- [22] HASSAN A, MAHMOUD M, AL-MAJED A, et al. A new technique to quantify the productivity of complex wells using artificial intelligence tools[C]. Dhahran: International Petroleum Technology Conference, 2020.
- [23] 张亦知, 程诚, 范钺彤, 等. 基于物理知识约束的数据驱动式湍流模型修正及槽道湍流计算验证[J]. *航空学报*, 2020, 41(3): 114-123.  
ZHANG Yizhi, CHENG Cheng, FAN Yitong, et al. Data-driven correction of turbulence model with physics knowledge constrains in channel flow[J]. *Acta Aeronautica et Astronautica Sinica*, 2020, 41(3): 114-123.
- [24] 朱国金, 田冀, 焦红梅, 等. 海上油田油井初期产能预测方法: Q/HS 2046—2014[S]. 北京: 中国海洋石油总公司, 2014.  
ZHU Guojin, TIAN Ji, JIAO Hongmei, et al. Predicting methods of well productivity in early period for offshore oil fields: Q/HS 2046-2014[S]. Beijing: China National Offshore Oil Corporation, 2014.
- [25] 刘浩男, 文晓涛, 何健, 等. 基于随机森林算法的AVO类型判别[J]. *中国海上油气*, 2020, 32(5): 73-81.  
LIU Haonan, WEN Xiaotao, HE Jian, et al. AVO type discrimination based on random forest algorithm[J]. *China Offshore Oil and Gas*, 2020, 32(5): 73-81.
- [26] 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.
- [27] 连克强. 基于Boosting的集成树算法研究与分析[D]. 北京: 中国地质大学(北京), 2018.  
LIAN Keqiang. The study and application of ensemble of trees based on Boosting[D]. Beijing: China University of Geosciences (Beijing), 2018.
- [28] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system[R]. San Francisco: International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.