

基于XGBoost算法的吸水剖面预测方法研究与应用

翟亮^{1,2}

(1. 中国石油大学(华东)石油工程学院, 山东 青岛 266580; 2. 中国石化胜利油田分公司勘探开发研究院, 山东 东营 257015)

摘要:胜利海上埕岛油田22F井区地质情况复杂,储层平面和纵向非均质性严重,制约了油田的注水开发效率。准确把握油藏各层间的注水情况进行油藏治理的重要前提,对于编制合理的注水开发方案也具有重要的指导意义。提出一种基于数据驱动的吸水剖面预测方法,利用Extreme Gradient Boosting(XGBoost)算法建立吸水剖面预测模型,根据油藏的地质参数和动态生产资料预测各注水井在整个开发时段内的吸水剖面演化规律,从而为合理生产配置和注采方案调整提供高质量的基础数据。在埕岛油田22F井区的应用结果表明,基于XGBoost算法建立的吸水剖面预测方法能够实现吸水剖面的准确反演和预测,平均相对误差为0.04,决定系数为0.87,均方根误差为3.12。与KH劈分方法相比,模型预测值与实际吸水量的吻合度更高,更能反映油藏的实际吸水情况,为油田的精细分层注水和智能开发夯实了基础,提供了技术支撑。

关键词:吸水剖面预测;数据驱动;机器学习;XGBoost;埕岛油田

中图分类号:TE319

文献标识码:A

XGBoost-based water injection profile prediction method and its application

ZHAI Liang^{1,2}

(1. School of Petroleum Engineering, China University of Petroleum (East China), Qingdao City, Shandong Province, 266580, China; 2. Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying City, Shandong Province, 257015, China)

Abstract: The complicated geological conditions and the strong horizontal and vertical heterogeneity of the reservoir severely restrict the water injection development efficiency of 22F Well Block in offshore Chengdao Oilfield, Shengli Oil Province. Accurately identifying the water injection situation of layers in a reservoir is an important prerequisite for reservoir management. It also has important guiding significance for the formulation of a reasonable water injection development plan. Therefore, a data-driven method is proposed for water injection profile prediction in this paper. The Extreme Gradient Boosting (XGBoost) algorithm is used to construct a model for making water injection profile prediction, with which the evolution of the water injection profile of each water injection well during the entire development period is predicted using the geological parameters and dynamic production data of the reservoir. As a result, high-quality data can be provided for rational production allocation and injection-production scheme adjustments. The application results in 22F Well Block of offshore Chengdao Oilfield show that the proposed method can accurately inverse and predict the water injection profile with an average relative error of 0.04, a determination coefficient of 0.87, and a root mean squared error of 3.12. Compared with the KH splitting method, the model in this paper yields a predicted value more consistent with the actual water absorption. This demonstrates that the proposed method can better reflect the actual water absorption of the reservoir and lays a solid foundation for fine stratified water injection and intelligent development of oilfields.

Key words: water injection profile prediction; data-driven; machine learning; XGBoost; Chengdao Oilfield

收稿日期:2021-08-12。

作者简介:翟亮(1978—),男,山东昌邑人,高级工程师,在读博士研究生,从事滩海油田开发研究工作。E-mail:zhailiang.slyt@sinopec.com。

基金项目:中国石化科技攻关项目“埕岛油田整体注采关键技术研究”(P17031-1)。

胜利油区埕岛油田22F井区是中国最早投入开发的年产油量百万吨级的滩海边际油田,位于渤海湾南部的极浅海海域,为典型的复式油气田,含油面积为1.27 km²,上报石油地质储量为5.29×10⁶ t,综合含水率为80.2%,采出程度为29.2%,处于高含水的初期阶段,剩余可采储量规模较大。为进一步提高油藏采收率,改善油田的整体开发效果,亟需开展吸水剖面的预测研究,准确把握油藏各个层段的注水情况^[1-3]。由于油藏纵向非均质性较严重,现有的KH劈分和劈分系数等方法未考虑注采系统和动态生产参数对注水井吸水剖面的影响,其劈分结果均不能准确反映油藏的实际吸水情况^[4-8]。

近年来,机器学习和人工智能技术在互联网^[9]、金融^[10]和智慧医疗^[11-12]等领域起到了推动作用,极大地提高了生产效率。目前,机器学习在石油行业的应用主要集中在钻完井^[13-14]、地质解释^[15]、测井解释^[16-17]和故障诊断^[18-20]等方面,在油藏工程上的研究相对较少。由于机器学习方法能够利用学习任务的数据,学习并生成特定于该任务的预测模型,挖掘数据中物理规律和知识,目前已在油气田开发中取得较好应用效果,主要体现在油井生产指标预测、井间连通性分析、生产优化和历史拟合等方面^[21-31]。基于此,利用吸水剖面资料,分析并确定影响吸水剖面的主要地质参数和动态生产参数,进而结合XGBoost算法^[32-33]建立吸水剖面预测模型,实现对吸水剖面的实时预测,为准确认识油藏各层段的注水情况提供数据基础,也为合理调配注水方法提供指导。

1 原理与方法

1.1 灰色关联分析

吸水剖面的动态变化主要受油藏地质参数和注采系统动态生产参数的影响,因此,利用回溯关联分析,对小层吸水量与注采系统中相关静态地质参数和动态开发参数进行关联性分析,依据关联度的大小筛选、确定影响吸水剖面的主要因素。关联度的计算包括4个步骤。

确定分析序列 吸水剖面的影响因素主要包括注采系统的静态地质参数和动态开发参数,静态地质参数主要包括孔隙度、渗透率和有效厚度等,动态开发参数包括产量、注入量和井口压力等。将影响吸水剖面的各个特征参数当作比较序列,构建分析矩阵为:

$$(X_1, X_2, \dots, X_n) = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(m) & x_2(m) & \cdots & x_n(m) \end{pmatrix} \quad (1)$$

以小层吸水量作为参考数据列 关联性分析目的在于确定与吸水剖面相关性较大的参数,因而以小层吸水量数据列作为参考数据列,表示为:

$$X_0 = (x_0(1), x_0(2), \dots, x_0(m)) \quad (2)$$

无量纲化序列 考虑到各序列参数之间的数量级和量纲的不同会影响关联性分析结果,因此需要对这些原始数据进行无量纲化处理,这里采用均值法对序列进行无量纲化处理,具体计算式为:

$$X'_i(k) = \frac{X_i(k)}{\frac{1}{n} \sum_{k=1}^n X_i(k)} \quad (3)$$

$$i = 0, 1, 2, \dots, n; k = 1, 2, \dots, m$$

计算吸水量与各因素之间的关联系数 依次计算由特征参数构成的比较序列与小层吸水量构成的参考数据列之间的关联度,即:

$$r_{0i} = \frac{1}{m} \sum_{k=1}^m \zeta_i(k) \quad (4)$$

其中:

$$\zeta_i(k) = \frac{\min_i^n |x_0(k) - x_i(k)| + \rho \max_{i=1}^n |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i=1}^n |x_0(k) - x_i(k)|} \quad (5)$$

依据各个影响因素与小层吸水量的关联度进行排序,从而剔除与小层吸水量关联度小的无关特征,构建最相关的特征参数,用于XGBoost模型的训练和预测。

1.2 XGBoost算法原理

XGBoost算法是一种以CART决策树模型为基础的集成学习方法,通过构建多棵CART决策树来提供预测模型的准确性,最后将每轮训练得到决策树的预测结果求和得到最终的预测值^[33]。相较于GBDT算法,XGBoost算法在目标函数中添加正则项,可有效防止过拟合,其第*t*棵CART决策树的目标函数定义如下:

$$L^{(t)} = \sum_{i=1}^n l(y_i, y_i^{(t-1)} + f_i(x_i)) + \Omega(f_i) \quad (6)$$

其中:

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (7)$$

在进行特征节点选择时,遍历训练集所有特征变量的取值,用节点分裂前的目标函数值减去分裂

后2个叶子节点的目标函数值之和,计算增益值,得到树模型最优的切分点,其中增益值计算式为:

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

1.3 模型预测效果评价

在测试集上评价预测模型的泛化能力,即模型的吸水剖面预测效果。采用的预测效果评价指标主要包括:决定系数 R^2 、平均相对误差 $MAPE$ 、均方误差 $RMSE$,其表达式分别为:

$$R^2 = 1 - \frac{\sum_{i=1}^N [\varepsilon_i - \hat{f}(x_i)]^2}{\sum_{i=1}^N (\varepsilon_i - \bar{\varepsilon})^2} \quad (9)$$

$$MAPE = \frac{1}{N} \cdot \sum_{i=1}^N \frac{|\varepsilon_i - \hat{f}(x_i)|}{\varepsilon_i} \quad (10)$$

$$RMSE = \left\{ \frac{1}{N} \sum_{i=1}^N [\varepsilon_i - \hat{f}(x_i)]^2 \right\}^{\frac{1}{2}} \quad (11)$$

2 实例分析

2.1 数据准备

考虑到吸水剖面预测模型的现场实用性,结合渗流理论和油藏工程师的经验,筛选出现场广泛易获取的参数作为初始的特征参数,对于注水井,主要包括:孔隙度、渗透率、有效厚度、注入压力、日注入水平和累积注水量;对于生产井,主要包括:孔隙度、渗透率、有效厚度、井距、日产液水平、油压、累积产油量、累积产水量和含水率。结合埕岛油田22F井区监测的吸水剖面资料,建立初始的吸水剖面数据集,部分数据如表1所示。

2.2 基于油藏工程方法的特征参数预处理

为了提高学习模型的准确性,利用油藏工程方法对选取的特征参数进行预处理。

注水井的注入压力 为了充分体现每个小层特征的差异性,需根据井口压力计算每个小层的注入压力,具体计算公式为:

$$p_i = p_1 + \rho_w g H_i \quad (12)$$

生产井的孔隙度和渗透率 生产井中小层的孔喉性质和渗流特性对注水井的吸水剖面产生影响,因而根据小层的厚度加权求平均孔隙度和平均

表1 部分吸水剖面数据
Table1 Water injection profile data

| 样本编号 | 孔隙度/% | 渗透率/mD | 有效厚度/m | 日注入水平/(m ³ ·d ⁻¹) | 日产液水平/(m ³ ·d ⁻¹) |
|------|--------|----------|--------|--|--|
| 1 | 35.57 | 2 394.92 | 7.49 | 170 | 70.2 |
| 2 | 32.92 | 963.39 | 1.46 | 170 | 70.2 |
| 3 | 33.22 | 242.84 | 1.80 | 170 | 70.2 |
| 4 | 35.03 | 1 576.29 | 4.99 | 170 | 70.2 |
| 5 | 38.26 | 3 255.52 | 8.74 | 170 | 70.2 |
| 6 | 35.56 | 2 394.92 | 7.49 | 141.7 | 100.4 |
| 7 | 32.92 | 963.39 | 1.46 | 141.7 | 100.4 |
| 8 | 33.22 | 242.84 | 1.80 | 141.7 | 100.4 |
| 9 | 35.03 | 1 576.29 | 4.99 | 141.7 | 100.4 |
| 10 | 38.26 | 3 255.52 | 8.74 | 141.7 | 100.4 |
| 11 | 35.57 | 2 394.92 | 7.49 | 116.1 | 130.6 |
| 12 | 32.92 | 963.39 | 1.46 | 116.1 | 130.6 |
| 13 | 33.22 | 242.84 | 1.80 | 116.1 | 130.6 |
| 14 | 35.033 | 1 576.29 | 4.99 | 116.1 | 130.6 |

渗透率。计算公式分别为:

$$\bar{\phi} = \frac{\sum_{i=1}^{I_y} \phi_i \sigma_i}{\sum_{i=1}^{I_y} \sigma_i} \quad (13)$$

$$\bar{K} = \frac{\sum_{i=1}^{I_y} K_i \sigma_i}{\sum_{i=1}^{I_y} \sigma_i} \quad (14)$$

2.3 特征参数优选

通过灰色关联分析,从众多影响因素中筛选出与吸水剖面最相关的特征,建立预测模型。各特征参数与吸水剖面的关联度分析结果如图1所示。

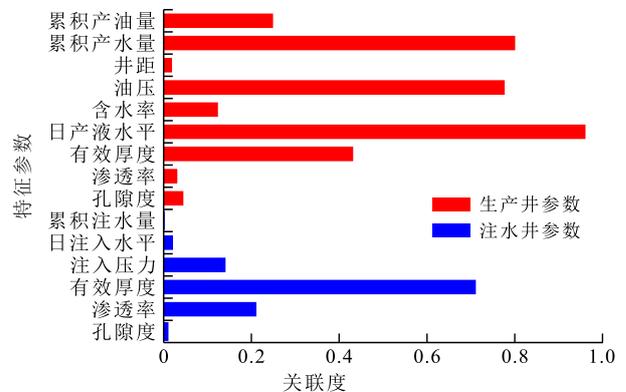


图1 各特征参数与吸水剖面的关联度

Fig.1 Correlations of characteristic parameters with water injection profile

从图1可以看出,注水井的有效厚度,生产井的日产液水平、油压、累积产水量与吸水剖面的相关性最强,关联度均大于0.7。注水井的渗透率、注入

压力,生产井的有效厚度、含水率、累积产油量与吸水剖面关联度为0.10~0.45,为中等强度的关联性,而其余参数与吸水剖面的关联度几乎为0,为弱关联或无关联特征。因此,选择与吸水剖面相关的特征参数:注水井的有效厚度、渗透率、注入压力和生产井的日产液水平、累积产水量、油压、有效厚度、累积产油量、含水率作为XGBoost模型的输入特征参数。结合特征属性和吸水剖面数据构成数据集,并以8:1:1的比例将该数据集划分为训练集、验证集和测试集,分别用于XGBoost模型的训练、参数调优和预测效果测试。

2.4 吸水剖面预测模型的训练和评价

XGBoost算法需要调优的模型超参数主要包括:学习率、树模型个数、最大树深度、正则参数 γ 、 α 和 λ 。采用K折交叉验证和网格搜索相结合的参数调优方法,确定XGBoost的最优超参数主要包括:学习率为0.15,树模型个数为25,最大树深度为7,正则参数 γ 为0.05, α 为0.03, λ 为1。

利用训练好的XGBoost模型预测测试集中的吸水剖面,计算各个模型评价指标分别为 R^2 为0.87, $MAPE$ 为0.96, $RMSE$ 为3.12。从模型在测试集上的评价指标可以看出,模型的预测值和实际值的吻合度较高,预测效果较好。选取测试集中注水井22FC-3的吸水剖面,应用XGBoost模型预测方法与KH劈分方法,由对比结果(图2)可以看出,XGBoost模型的预测值与实际值的吻合度要好于KH劈分方法,说明考虑了动态参数对吸水剖面影响的XGBoost预测模型更能反映油藏的实际吸水情况,满足矿场的应用需求。

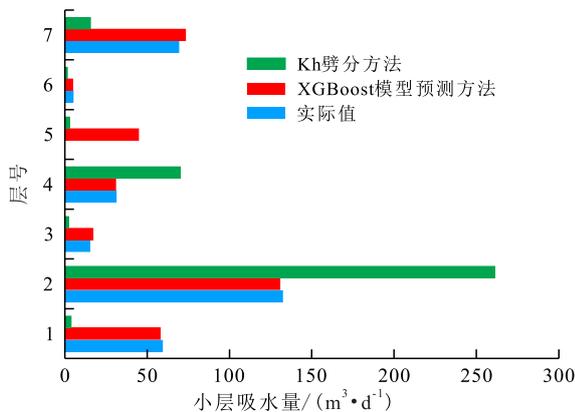


图2 XGBoost模型预测值与实际吸水剖面的对比结果

Fig.2 Comparison results of XGBoost model prediction with actual water injection profile

3 结论

提出了基于XGBoost算法的吸水剖面预测方

法,能够利用井区现有的吸水剖面资料,通过训练和学习得到吸水剖面预测模型,可有效反演和预测吸水剖面的动态变化。通过灰色关联分析准确提取了与吸水剖面最相关的静态地质参数和动态开发参数,用于构建预测模型的输入特征,有助于认识吸水剖面的影响机理。同时,模型的复杂度得以有效降低,模型的训练速度和泛化性能得以提高。利用XGBoost模型准确挖掘了吸水剖面与各参数之间的相关关系,能够实现对吸水剖面的连续准确预测。在埕岛油田22F井区的应用研究表明,所提出的吸水剖面预测方法相较于KH劈分方法,预测效果更好,便于油田实际应用,具有较高的矿场应用价值。

符号解释

- $f_t(x_i)$ ——第 t 棵CART决策树的预测值;
- $\hat{f}(x_i)$ ——XGBoost模型在 i 时刻或第 i 个样本处的小层吸水量预测值, m^3/d ;
- g ——重力加速度, m/s^2 ,取值为9.8;
- g_t ——损失函数的一阶导数;
- h_t ——损失函数的二阶导数;
- i ——样本索引号;
- H_t ——小层 i 的油藏中深, m ;
- I ——节点分裂前的总样本集;
- I_L ——节点分裂后的左子集;
- I_R ——节点分裂后的右子集;
- k ——特征变量取值索引号;
- \bar{K} ——生产井的平均渗透率, mD ;
- K_t ——小层 i 的渗透率, mD ;
- l ——损失函数,用于衡量预测值与真实值间的误差;
- l_y ——生产井的小层数;
- $L^{(t)}$ ——第 t 棵CART决策树的目标函数;
- L_{split} ——节点分裂的增益值;
- m ——影响因素的取值个数或维度;
- $MAPE$ ——平均相对误差;
- n ——初步确定的吸水面影响因素数量,个;
- N ——样本数,个;
- p_i ——小层 i 的注入压力, MPa ;
- p_l ——注水井的井口压力, MPa ;
- r_{oi} ——第 i 个影响因素与小层吸水量之间的关联度,小数;
- R^2 ——决定系数;
- $RMSE$ ——均方误差;
- t ——决策树索引号;
- T ——第 t 棵CART决策树的叶子节点数;
- w ——第 t 棵CART决策树的叶子节点权重;
- $x_0(m)$ ——参考数据列特征变量的第 m 个取值,为实数;

x_n ——第 n 个特征变量的取值,为实数;
 $x_n(m)$ ——第 n 个特征变量的第 m 个取值,为实数;
 X_0 ——参考数据列;
 $X_i(k)$ ——第 i 个特征变量的第 k 个取值,为实数;
 $X'_i(k)$ ——第 i 个无量纲化特征变量的第 k 个取值,为实数;
 X_n ——第 n 个特征变量;
 y_i ——研究目标的真实值;
 $y'_i(t-1)$ ——第 $t-1$ 棵 CART 决策树的预测值;
 α, λ ——第 t 棵 CART 决策树的叶子权重惩罚正则项系数;
 γ ——第 t 棵 CART 决策树的叶子节点惩罚正则项系数;
 $\bar{\varepsilon}$ ——实测小层吸水量的平均值, m^3/d ;
 ε_i ——第 i 个样本或第 i 时刻的实测小层吸水量, m^3/d ;
 ρ ——分辨系数,其值为 $0 \sim 1$,其值越小,关联系数间差异越大,区分能力越强,通常取值为 0.5 ;
 ρ_w ——水的密度, kg/m^3 ;
 σ_i ——小层 i 的厚度, m ;
 $\bar{\phi}$ ——生产井的平均孔隙度,小数;
 ϕ_i ——小层 i 的孔隙度,小数;
 $\Omega(f_i)$ ——第 t 棵 CART 决策树的正则项。

参考文献

- [1] 袁士义,王强.中国油田开发主体技术新进展与展望[J].石油勘探与开发,2018,45(4):657-668.
YUAN Shiyi, WANG Qiang. New progress and prospect of oil fields development technologies in China[J]. Petroleum Exploration and Development, 2018, 45(4): 657-668.
- [2] 杜庆龙,朱丽红.油、水井分层动用状况研究新方法[J].石油勘探与开发,2004,31(5):96-98.
DU Qinglong, ZHU Lihong. A new approach to study layered producing performance of oil and water wells[J]. Petroleum Exploration and Development, 2004, 31(5): 96-98.
- [3] 杜庆龙.多层非均质砂岩油田小层动用状况快速定量评价方法[J].大庆石油地质与开发,2016,35(4):43-48.
DU Qinglong. Quick-quantitative evaluating method of the sublayer developed conditions in the multilayered heterogeneous sandstone oilfield[J]. Petroleum Geology & Oilfield Development in Daqing, 2016, 35(4): 43-48.
- [4] 谢军,于家义,张中劲,等.油田开发井组产层分层动用新方法研究[J].石油天然气学报,2017,39(5):238-244.
XIE Jun, YU Jiayi, ZHANG Zhongjin, et al. The new method for separating layer production in oilfield development well group[J]. Journal of Oil and Gas Technology, 2017, 39(5): 238-244.
- [5] 林承焰,孙廷彬,董春梅,等.基于单砂体的特高含水期剩余油精细表征[J].石油学报,2013,34(6):1 131-1 136.
LIN Chengyan, SUN Tingbin, DONG Chunmei, et al. Fine characterization of remaining oil based on a single sand body in the high water cut period[J]. Acta Petrolei Sinica, 2013, 34(6): 1 131-1 136.
- [6] 于洪文.通过注水井分层注水量计算采油井分层产液(油)量的方法探讨[J].石油勘探与开发,1990,45(2):47-54.
YU Hongwen. A discussion using the techniques of production profile determination by the injection profiles of corresponding injectors[J]. Petroleum Exploration and Development, 1990, 45(2): 47-54.
- [7] 熊昕东,杨建军,刘坤,等.运用劈分系数法确定注水井单井配注[J].断块油气田,2004,11(3):56-59.
XIONG Xindong, YANG Jianjun, LIU Kun, et al. Determination of water injection rate for single injection well employing dividing coefficient approach[J]. Fault-Block Oil and Gas Field, 2004, 11(3): 56-59.
- [8] 黄学峰,李敬功,吴长虹,等.注水井分层累计吸水量动态劈分方法[J].测井技术,2004,28(5):465-467.
HUANG Xuefeng, LI Jingong, WU Changhong, et al. Dynamic division method of accumulated injected-water volume for water-injected well[J]. Well Logging Technology, 2004, 28(5): 465-467.
- [9] 李业志,陈艳,胡悦.大数据在互联网经济发展中的应用[J].计算机光盘软件与应用,2014,17(8):89,91.
LI Yezhi, CHEN Yan, HU Yue. The application of big data in the development of internet economy[J]. Computer CD Software and Applications, 2014, 17(8): 89, 91.
- [10] 米传民,徐润捷,陶静.互联网金融空间聚集分析及系统性风险防范——基于t-SNE机器学习模型[J].财经论丛,2019,52(8):53-62.
MI Chuanmin, XU Runjie, TAO Jing. Analysis of internet financial spatial aggregation and systematic risk prevention-based on t-SNE machine learning model[J]. Collected Essays on Finance and Economics, 2019, 52(8): 53-62.
- [11] 杨薇,崔英子,杨海森,等.医疗大数据在中医药研究领域的应用与思考[J].长春中医药大学学报,2016,32(3):625-627.
YANG Wei, CUI Yingzi, YANG Haimiao, et al. Thinking of medical big data applied to Chinese medical research[J]. Journal of Changchun University of Chinese Medicine, 2016, 32(3): 625-627.
- [12] 李蓉,孙媛.机器学习技术在胸癌诊断中的应用[J].科学技术与工程,2011,11(20):4 730-4 733,4 739.
LI Rong, SUN Yuan. Breast cancer diagnosis using machine learning technique[J]. Science Technology and Engineering, 2011, 11(20): 4 730-4 733, 4 739.
- [13] 梁文福.油田开发智能应用系统建设成果及展望[J].大庆石油地质与开发,2019,38(5):283-289.
LIANG Wenfu. Constructed achievements and prospects of the intelligent application system for the oilfield development[J]. Petroleum Geology & Oilfield Development in Daqing, 2019, 38(5): 283-289.
- [14] 耿黎东.大数据技术在石油工程中的应用现状与发展建议[J].石油勘探技术,2021,49(2):72-78.
GENG Lidong. Application status and development suggestions of big data technology in petroleum engineering[J]. Petroleum Drilling Techniques, 2021, 49(2): 72-78.
- [15] 赵军,程鹏飞,刘地渊,等.支持向量机在水淹层测井识别中的应用[J].物探与化探,2008,32(6):652-655.

- ZHAO Jun, CHENG Pengfei, LIU Diyan, et al. The application of the support vector machine to the recognition of flooding formation [J]. *Geophysical and Geochemical Exploration*, 2008, 32(6): 652-655.
- [16] 石玉江, 刘国强, 钟吉彬, 等. 基于大数据的测井智能解释系统开发与应用[J]. *中国石油勘探*, 2021, 26(2): 113-126.
- SHI Yujiang, LIU Guoqiang, ZHONG Jibin, et al. Development and application of intelligent logging interpretation system based on big data [J]. *China Petroleum Exploration*, 2021, 26(2): 113-126.
- [17] 李政宏, 刘永福, 张立强, 等. 数据挖掘方法在测井岩性识别中的应用[J]. *断块油气田*, 2019, 26(6): 713-718.
- LI Zhenghong, LIU Yongfu, ZHANG Liqiang, et al. Application of data mining method in lithology identification using well log [J]. *Fault-Block Oil and Gas Field*, 2019, 26(6): 713-718.
- [18] 舒服华. 基于加权支持向量机的抽油机井故障诊断[J]. *石油机械*, 2006, 34(3): 57-59.
- SHU Fuhua. Fault diagnosis of pumping well based on weighted support vector machine [J]. *China Petroleum Machinery*, 2006, 34(3): 57-59.
- [19] 孙锦辉. 石油生产中的故障诊断与产品质量预测[D]. 北京: 北京邮电大学, 2019.
- SUN Jinhui. Fault diagnosis and product quality prediction in petroleum production [D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [20] 陈国明, 朱高庚, 朱渊. 深水油气开采安全风险评估与管控研究进展[J]. *中国石油大学学报: 自然科学版*, 2019, 43(5): 136-145.
- CHEN Guoming, ZHU Gaogeng, ZHU Yuan. Advances in safety assessment and risk management for deepwater oil and gas exploitation [J]. *Journal of China University of Petroleum: Edition of Natural Science*, 2019, 43(5): 136-145.
- [21] 任燕龙, 谷建伟, 崔文富, 等. 基于改进果蝇算法和长短期记忆神经网络的油田产量预测模型[J]. *科学技术与工程*, 2020, 20(18): 7245-7251.
- REN Yanlong, GU Jianwei, CUI Wenfu, et al. Oilfield production prediction model based on improved fruit fly algorithm and long-short term memory neural network [J]. *Science Technology and Engineering*, 2020, 20(18): 7245-7251.
- [22] 谷建伟, 隋顾磊, 李志涛, 等. 基于ARIMA-Kalman滤波器数据挖掘模型的油井产量预测[J]. *深圳大学学报: 理工版*, 2018, 35(6): 575-581.
- GU Jianwei, SUI Gulei, LI Zhitao, et al. Oil well production forecasting method based on ARIMA-Kalman filter data mining model [J]. *Journal of Shenzhen University: Science & Engineering*, 2018, 35(6): 575-581.
- [23] 刘巍, 刘威, 谷建伟. 基于机器学习方法的油井日产油量预测[J]. *石油钻采工艺*, 2020, 42(1): 70-75.
- LIU Wei, LIU Wei, GU Jianwei. Oil production prediction based on a machine learning method [J]. *Oil Drilling & Production Technology*, 2020, 42(1): 70-75.
- [24] LIU W, LIU W D, GU J, et al. Predictive model for water absorption in sublayers using a machine learning method [J]. *Journal of Petroleum Science and Engineering*, 2019, 182: 106367.
- [25] 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.
- [26] 谷建伟, 任燕龙, 王依科, 等. 基于机器学习的平面剩余油分布预测方法[J]. *中国石油大学学报: 自然科学版*, 2020, 44(4): 39-46.
- GU Jianwei, REN Yanlong, WANG Yike, et al. Prediction methods of remaining oil plane distribution based on machine learning [J]. *Journal of China University of Petroleum: Edition of Natural Science*, 2020, 44(4): 39-46.
- [27] 但松林, 刘尚奇, 罗艳艳, 等. 基于BP神经网络预测高含水层对SAGD开发效果的影响[J]. *大庆石油地质与开发*, 2019, 38(2): 73-80.
- DAN Songlin, LIU Shangqi, LUO Yanyan, et al. Predicted SAGD development effects by BP neural network for the high-watercut reservoir [J]. *Petroleum Geology & Oilfield Development in Daqing*, 2019, 38(2): 73-80.
- [28] 陈雁, 焦世祥, 程超, 等. 基于自编码器的半监督隔夹层识别方法[J]. *特种油气藏*, 2021, 28(1): 86-91.
- CHEN Yan, JIAO Shixiang, CHENG Chao, et al. Semi-supervised interlayer identification method based on self-encoder [J]. *Special Oil & Gas Reservoirs*, 2021, 28(1): 86-91.
- [29] 程翊珊, 李治平, 许龙飞, 等. 预测油层无机积垢的BP神经网络方法[J]. *大庆石油地质与开发*, 2021, 40(3): 84-93.
- CHENG Yishan, LI Zhiping, XU Longfei, et al. BP neural network method for predicting the inorganic scaling in the reservoir [J]. *Petroleum Geology & Oilfield Development in Daqing*, 2021, 40(3): 84-93.
- [30] 黄家宸, 张金川. 机器学习预测油气产量现状[J]. *油气藏评价与开发*, 2021, 11(4): 613-620.
- HUANG Jiachen, ZHANG Jinchuan. Overview of oil and gas production forecasting by machine learning [J]. *Reservoir Evaluation and Development*, 2021, 11(4): 613-620.
- [31] 纪磊, 李菊花, 肖佳林. 随机森林算法在页岩气田多段压裂改造中的应用[J]. *大庆石油地质与开发*, 2020, 39(6): 168-174.
- Ji Lei, LI Juhua, XIAO Jialin. Application of random forest algorithm in the multistage fracturing stimulation of shale gas field [J]. *Petroleum Geology & Oilfield Development in Daqing*, 2020, 39(6): 168-174.
- [32] ANG J C, MIRZAL A, HARON H, et al. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2015, 13(5): 971-989.
- [33] CHEN T, GUESTIN C. Xgboost: A scalable tree boosting system [C]. San Francisco: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.