

基于机器学习的油藏生产监测数据分析与预警方法

李建荣, 王胜

中国石油化工股份有限公司胜利油田分公司 山东东营 257001

通讯作者: Email: wangsheng658. slyt@ sinopec. com

项目支持: 中石化重大先导课题“水平井动态监测技术研究与应”(P22024); 胜利油田分公司油藏动态监测中心科研“基于监测资料的区块开发动态分析与预测技术”(ZX2022-10)

引用: 李建荣, 王胜. 基于机器学习的油藏生产监测数据分析与预警方法[J]. 油气井测试, 2023, 32(5): 8-13.

Cite: LI Jianrong, WANG Sheng. Reservoir production monitoring data analysis and early warning method based on machine learning[J]. Well Testing, 2023, 32(5): 8-13.

摘要 对于渗流机理不明确或地质模型不确定性高的非常规油气藏, 基于油藏工程和数值模拟法的油藏动态分析误差较大。利用生产监测数据, 分析油藏动态, 实现数据驱动的监测井网优化和生产动态预测预警。采用 K 均值法建立生产井聚类算法, 实现生产监测数据驱动的监测井网优化; 构建卷积神经网络拟合生产监测指标间非线性关系, 量化指标的不确定性区间, 通过计算小概率事件进行预警。经实际算例验证表明: 监测井网优化算法能够有针对性地在特征空间实现聚类, 得到最具代表性的井位构成监测井网; 指标预测预警方法能够稳健地预测指标不确定性区间, 作为预警的依据。该方法为监测井网优化和生产动态预警提供了一种不依赖渗流机理的数据驱动技术, 对智能油气田建设有积极作用。

关键词 动态监测; 数据驱动; 井网优化; K 均值聚类; 卷积神经网络; 指标预测; 动态预警; 机器学习

中图分类号: TE353

文献标识码: A

DOI: 10. 19680/j. cnki. 1004-4388. 2023. 05. 002

Reservoir production monitoring data analysis and early warning method based on machine learning

LI Jianrong, WANG Sheng

Shengli Oilfield, Sinopec, Dongying, Shandong 257001, China

Abstract: For unconventional oil and gas reservoirs with unclear flow mechanisms or high geological model uncertainties, traditional reservoir engineering and numerical simulation methods can introduce significant errors in dynamic reservoir analysis. To address this issue, by using production monitoring data, the reservoir dynamics was analyzed to realize data-driven monitoring well network optimization, as well as dynamic production forecasting and early warning. By using a K-means method, a clustering algorithm for production wells was established, which achieved monitoring well network optimization driven by production monitoring data. Additionally, a convolutional neural network was constructed to model the nonlinear relationships between production monitoring indicators, as well as to quantify the uncertainty intervals of these indicators, and then, early warning was realized by calculating small probability events. Verified by actual cases, it is shown that the monitoring well network optimization algorithm can achieve targeted clustering in the feature space and obtain the most representative well locations to form a monitoring well network; the indicator prediction and early warning method can robustly predict the indicator uncertainty interval and serve as the basis for early warning. This approach provides a data-driven technique that does not rely on flow mechanisms and has a positive impact on the establishment of intelligent oil and gas fields, particularly for reservoir monitoring network optimization and dynamic production forecasting and early warning.

Keywords: dynamic monitoring; data driven; well network optimization; K-means clustering; convolutional neural network; indicator prediction; dynamic early warning; machine learning

动态监测技术经过多年发展逐步形成了基础实验、监测方案设计、监测资料采集、数据解释和综合应用等方面的技术, 成为油藏工程和油藏数值模拟之外认识与评价油藏开发动态和效果的重要手

段^[1-3]。基于渗流力学方程解析解的传统油藏工程方法需对油藏模型大幅简化, 无法充分考虑储层非均质性以及复杂流动机理的影响。油藏数值模拟是油藏动态分析和预测的常规方法, 其优点是参数

间基于机理的相关性明确,经过多年发展方法成熟且可基于商用软件实现;其主要局限性是依赖准确的地质模型和渗流机理,对于地质模型或渗流机理不明确的区块误差大,且只能考虑部分监测数据^[4-5]。数据驱动方法是利用机器学习和深度学习等方法直接拟合数据间非线性关系,并进行相应的分析与预测^[6-8],无需明确的渗流机理或地质模型,但对采集的数据质量要求较高^[9-11]。

基于机器学习的数据分析主要有相关性分析、去噪与降维、分类与回归、预测、反演与控制等方面,并已经形成了 K 近邻法、支持向量机、逻辑斯蒂回归和神经网络等算法。如何将机器学习算法与具体工程实际结合,建立相应的数据驱动分析技术是工业大数据研究的重点。针对油藏生产监测数据,基于机器学习建立相应的数据分析算法^[12-14],有利于充分挖掘监测数据特征,综合考虑数据各维度关系^[15-16],使监测资料在指导开发中发挥更大作用,对智能油气田建设具有重要意义。

油藏监测项目分为连续监测的生产指标(产油量、产液量、井底流压等)和定期监测的注入产出剖面、剩余油饱和度、地球物理测井和工程测井等。本文基于机器学习建立生产监测数据驱动的监测井聚类 and 监测指标预测预警算法^[17-18],为成本较高的定期监测项目高效实施提供依据。本文首先介绍实际生产监测数据的维度,进而对其分类进行相关性分析,实现降维。然后,基于 K 均值法针对第一步分类的数据进行聚类分析,并构建图版确定监测井位数目,得到监测井网作为定期监测的依据。进一步基于深度学习拟合指标间函数关系,并据此预测指标的不确定性区间,从而实现动态预警,预警的井位也需增加至监测井网中。

1 监测数据相关性分析

某典型岩性构造油藏主力油层埋藏深度大致在 1.5~1.6 km 之间,构造形态为在单斜背景上的一系列小型断块,主要发育三条大断裂及众多小断裂。地质模型三维结构不确定性高,储层物性非均质性强,导致地质建模困难,但油藏生产监测资料丰富。监测指标包括井口温度、井底温度、油嘴、井口压力、井底流压、产油量、产气量、产水量、产液量、含水率、泵频率、泵电流、泵电压、泵漏电电流、泵震动、泵入口温度、泵马达温度、泵入口压力、泵出口压力共 19 项。监测数据来自区块内 25 口井采

用电潜泵的生产井。将所有井同一指标的数据拼接至同一维度作为整个区块的数据。首先将数据分类为油藏数据(井口温度、井底温度、油嘴、井口压力、井底流压、产油量、产气量、产水量、产液量、含水率)与工程数据(泵频率、泵电流、泵电压、泵漏电电流、泵震动、泵入口温度、泵马达温度、泵入口压力、泵出口压力)。

根据油藏工程知识,产液量为产油量和产水量之和,含水率为产水量比产液量,因此,在油藏数据中忽略产液量和含水率,防止重复数据影响后续的聚类结果。进一步构建相关系数矩阵初步分析各指标间相关性。两个随机变量 X 和 Y 间的皮尔逊相关系数为

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (1)$$

式中: $Cov(X, Y)$ 为两者间协方差; σ_X 和 σ_Y 分别为两者标准差。

计算了所有 19 项监测指标间的相关系数,发现泵入口温度与井底温度强相关;泵入口压力与井底流压强相关。这些均与主观认知相符,可作为对数据的检验。由于泵马达温度与泵入口温度强相关(相关系数 0.962),泵频率与泵电压强相关(相关系数 0.907),因此从工程数据中根据因果关系忽略泵频率和泵马达温度。

2 监测数据聚类分析

油藏监测定期监测项目实施成本高,对每一口井进行动态监测工程量过大。通过 K 均值聚类法分别基于油藏数据和工程数据将 25 口井进行聚类,找到每一类中的最具代表性的井作为这一类的监测井,构成重点监测井网,以此提高监测效率。

2.1 K 均值聚类

K 均值聚类是无监督学习算法,即无需提前训练,可直接应用于实际数据。其计算复杂度接近线性,收敛速度快,可解释性强,但只能保证局部最优。其理论保证在于, K 均值聚类是使用最大期望算法求解的高斯混合模型在正态分布的协方差为单位矩阵,且隐变量的后验分布为一组狄拉克函数时所得到的特例。K 均值聚类对凸数据聚类效果好,且对噪声敏感。

K 均值聚类算法是一种迭代求解的聚类分析算法。其步骤是,将数据预分类为 K 组,随机选取 K 个对象作为初始的聚类中心,然后计算每个对象与各聚类中心之间的距离,把每个对象分配给距离它

最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本,聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有(或最小数目)对象被重新分配给不同的聚类,没有(或最小数目)聚类中心再发生变化,误差平方和即聚类误差局部最小。误差平方和(SSE)的定义如下

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \tag{2}$$

式中: C_i 为第 i 个类; p 为 C_i 中的井; m_i 为 C_i 的中心; SSE 为所有井位的聚类误差,代表聚类效果。

2.2 图版法确定监测井位数

K 均值聚类法唯一的超参数是 K 的值,对应监测井位数,这也是监测井网优化的关键。K 均值聚类法实质是在给定 K 的前提下将聚类误差最小化。对此,可构建聚类误差随监测井位数变化的图版,然后采用手肘法确定最佳聚类数目。随着聚类数 K 的增大,样本划分会更加精细,每个簇的聚合程度会逐渐提高,聚类误差也随之逐渐变小。当 K 小于真实聚类数时,K 的增大会大幅增加每个簇的聚合程度,故 SSE 的下降幅度会很大,而当 K 到达真实聚类数时,再增加 K 所得到的聚合程度回报会迅速变小,导致 SSE 的下降幅度骤减,从而 SSE 随 K 的变化图版呈现手肘的形状,而这个肘部对应的 K 值就是数据的真实聚类数。

2.3 监测井网聚类

根据所有油藏和工程数据,采用手肘法确定聚类数目,结果如图 1 所示。油藏数值模拟需要考虑储层物性、含油饱和度、原始油藏压力等,这些都是影响开发动态的关键指标。此处对井位进行聚类分析,依赖的是动态数据,即井口温度、井底温度、井口压力、井底流压等动态油藏数据以及泵的动态数据。从渗流力学方程角度,井底流压和产液量等动态数据间接反映了储层物性参数。根据图 1 曲线手肘处选择最小聚类数目为 3,在相关性分析后保留的特征空间(15 维)上进行聚类分析(见表 1)。将表 1 可视化,如图 2 所示,不同颜色代表不同类,每一聚类类中心最近的井位作为监测井,这样得到的监测井网最具代表性,即 7,5,17。也可以根据需求自由选择监测井的数目,以 K=5 为例同样在特征空间上进行聚类分析,确定监测井位为 7,17,18,10,20。

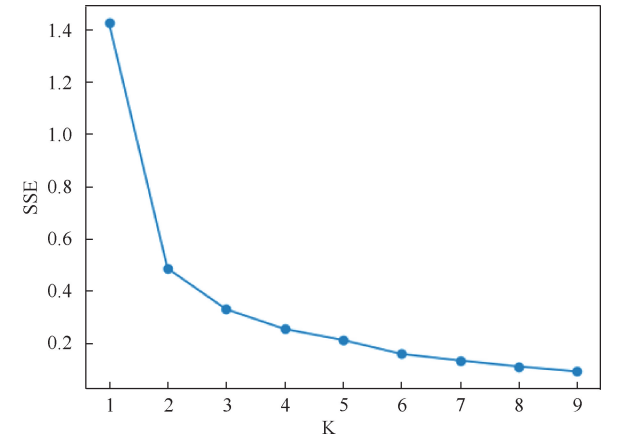


图 1 误差平方和(SSE)与聚类数目 K 关系图
Fig.1 The relationship between SSE and k for clustering

表 1 聚类结果(聚类数目 K=3)

Table 1 Clustering result (number of clusters k=K)		
聚类编号	类中包含井位	各井位距离类中心距离
A	1,4,7	2 816,6 449,1 436
B	5,10,11,14,21	1 827,2 795,5 239,5 782,5 548
C	2,3,6,8,9,12,13,15,16,17,18,	2 146,2 625,1 559,2 976,2 699,
	19,20,22,23,	5 618,1 921,1 125,4 519,785,
	24,25	8 360,6 871,3 998,2 617,1 308,3 905,2 233

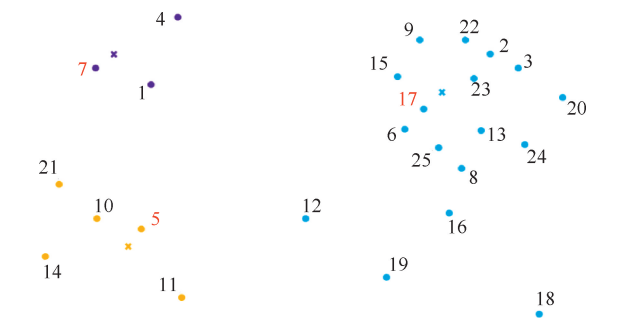


图 2 聚类结果可视化(聚类数目 K=3)
Fig.2 Visualization of clustering result (number of clusters K=3)

油藏数据和工程数据分别反映了生产井的地质和设备状态,基于两者进行聚类分析得到的监测井网更有针对性。基于油藏数据(井口温度,井底温度,油嘴,井口压力,井底流压,油,气,水)聚类分析,根据手肘法得到的监测井位为 1,5,7,17,20。基于工程数据(泵电流,泵电压,泵漏电电流,泵震动,泵入口温度,泵入口压力,泵出口压力)聚类分析,根据手肘法得到的监测井位为 6,12,13,14,15,18。然后分别对两组井位进行监测。综合上述数据相关性分析和聚类分析,形成生产监测数据分析和井网优化工作流程如图 3 所示。

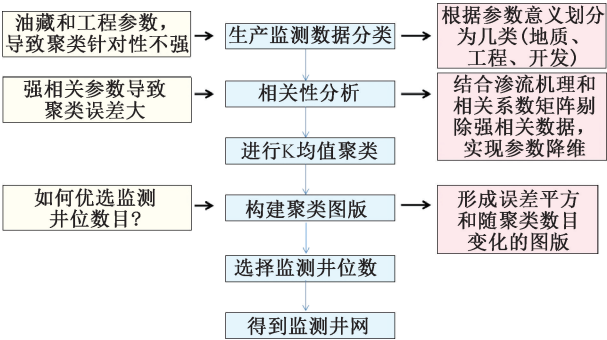


图 3 生产监测数据驱动的监测井网优化工作流程
Fig.3 Workflow of production monitoring data driven optimization of the monitoring well network

3 监测数据预测预警

在油田开发过程中不合理的开发方式和不健康的开发状态是影响油田长期高产、稳产目标的主要因素,也是根本性原因。尤其是在老油田进入开发中后期,油藏各种矛盾突出,开发方案调整的不及时、不合理,油水井带病运行不仅带来生产成本的大幅增加,同时对油藏带来不可逆的伤害和最终采收率的降低。

基于监测数据进行预测预警可以及时反映油藏生产动态是否正常。以井底温度为 例,首先根据相关性分析发现,井底流压和井底温度相关性高,可依据其中之一预测另一个指标,以此作为预警的依据。当实测数据与预测数据偏差较大时进行预警。

3.1 卷积神经网络(CNN)与预警原理

CNN 是一类具有深度结构的前馈神经网络,是深度学习的代表算法之一,在计算机视觉等领域应用广泛。卷积神经网络基于卷积核进行特征提取,抗噪效果好且比全连接神经网络更高效。因此,采用卷积神经网络作为算子拟合指标间非线性关系,作为监测指标预测模型。以井底流压为输入,井底温度最大与最小值为输出,采用 Adam 优化器进行训练得到两个神经算子,即

$$\tilde{T}_{\max} = f(p_{wf}) \tag{3}$$

$$\tilde{T}_{\min} = g(p_{wf}) \tag{4}$$

式中: $p_{wf} = p_{wf}(t)$ 为井底流压,MPa; $\tilde{T}_{\max} = \tilde{T}_{\max}(t)$ 为井底温度最大值的近似值; $\tilde{T}_{\min} = \tilde{T}_{\min}(t)$ 为井底温度最小值的近似值; f 为从井底流压 p_{wf} 到井底温度最大值近似值 \tilde{T}_{\max} 映射的神经算子; g 为从井底流压 p_{wf} 到井底温度最小值近似值 \tilde{T}_{\min} 映射的神经算子。

式(3)-式(4),即

$$\tilde{T}_{\max} - \tilde{T}_{\min} = (f - g)(p_{wf}) \tag{5}$$

式(5)对应某时刻井底流压值的井底温度变动区间,从而量化井底温度的不确定性。作为神经算子的深度卷积神经网络结构见图 4。

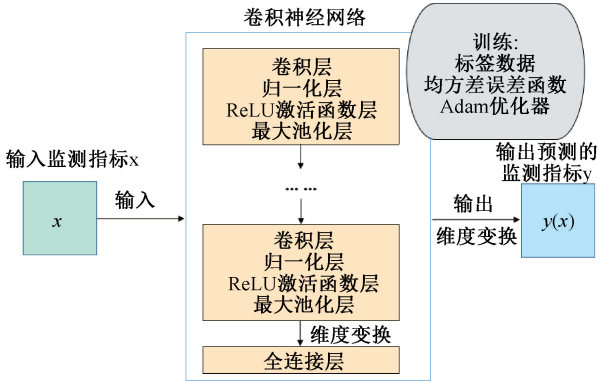


图 4 深度卷积神经网络结构
Fig.4 Network Structure of Deep CNN

利用卷积神经网络训练井底温度最值预测模型,当一口井工作时,输入测量得到的井底流压值,模型预测出相应的井底温度最大值 y_{\max} (最小值 y_{\min})。考虑到测量误差的影响,若测量所得的真实温度持续 $y \geq y_{\max}$ ($y \leq y_{\min}$) 则发出预警。预警的基础是小概率事件发生。因此,提出下述生产监测数据驱动的预警算法:

- (1) 基于历史数据构建测试集;
- (2) 利用预测模型得到预警指标的上下界;
- (3) 基于测试集计算落在不确定性区间外的概率即 $p = N_{\text{out}}/N$, 其中, N 为测试集总样本数, N_{out} 为落在不确定性区间外的样本数;
- (4) 确定小概率阈值 ε (比如 0.01%), 则 $n = \log_p \varepsilon$ 。从而,如果持续 n 次实测数据落在不确定性区间外,则进行预警。

3.2 井底温度最值预测模型

为了减少相邻井之间相关性较大所引起的系统误差,等间隔选取 5 口井(5, 10, 15, 20, 25 号井)作为测试集,将剩余 20 口井作为训练集。预测样本数为 9 196,将剩余 20 口井作为训练集,训练样本数为 21 823。单独抽取井底流压与井底温度指标,进行数据清洗(去除含 NA 的数据)后,按照井底流压分组,选取组内井底温度最大值与最小值,形成数据矩阵。观察发现,由于部分井底流压只对应一个数据点,无法反映真实的温度最值,故增大井底流压步长(0.01→0.05 kPa),以消除异常点对结果的影响。以井底流压作为输入,井

底温度最大值作为输出,得到井底温度最大值预测模型如图 5 所示。

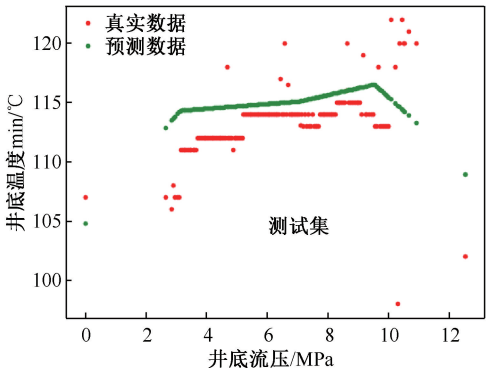
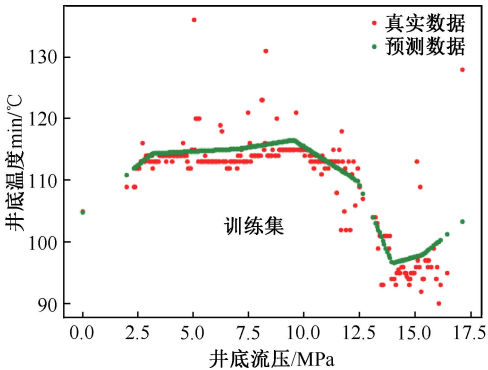


图 5 井底温度最大值预测结果
Fig. 5 Prediction results of maximum bottom-hole pressure

由图 5 分析可得,在训练集上数据拟合程度高,拟合标准差为 3.863 9。在测试集上预测温度最大值高于实际温度的比例为 0.964 8,这表明预测温度最大值以很大的比例涵盖实际测量温度,同时说明该模型能够较为稳定的进行预警。井底温度最小值预测模型同理,拟合的标准误差为 6.516 3,在测试集上预测温度最小值低于实际温度的比例为 0.954 7。综合两个模型得出,井底温度最值预测模型提供的最值区间至少以 95% 的概率包含实际测量温度,说明该模型能够较为稳定的进行预测预警,结果如图 5 所示。在预测集上,绿线是预测的

5 口井的最高井底温度。由于训练数据来自区块内的 20 口井,而测试集为区块内的 5 口井,训练集覆盖范围更广,因此一般情况下,基于训练数据预测的最高井底温度高于测试集的实际最高井底温度。当实际最高井底温度持续高于预测的井底温度时进行预警。同理如图 6 所示,预测的井底温度最低值也低于实际井底温度最低值。由于最值区间包含实际数据的概率为 95%,则不包括的概率仅为 5%,而连续 3 次测量数据落在不确定性区间外的几率仅为 0.000 125,为小概率事件。因此,当连续 3 次测量数据落在不确定性区间外时进行预警。

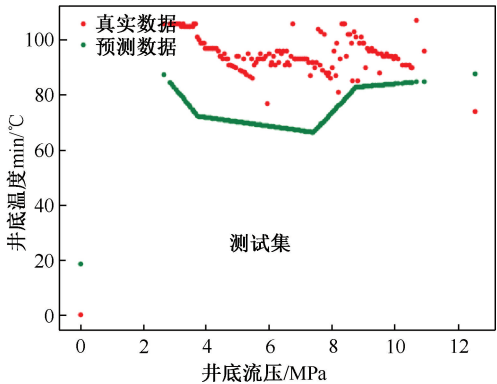
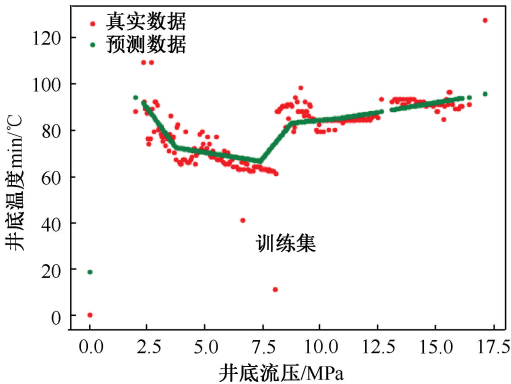


图 6 井底温度最小值预测结果
Fig. 6 Prediction results of minimum bottom-hole pressure

4 结论

(1)油藏工程和数值模拟应用于渗流机理不明确和地质模型不确定性高的非常规油气藏误差较大,相应的监测井网优化困难。以连续的生产监测数据为数据基础,以 K 均值聚类法为算法基础,建立了数据驱动的监测井网优化算法和 workflow,并应用于了真实生产监测数据,得到针对油藏数据和工程数据的监测井网,作为定期监测项目的选井

依据。通过建立聚类误差与聚类数目间关系模板确定监测井位数目。

(2)针对生产监测数据特点,建立了一种基于指标函数关系的预警方法,即通过深度学习拟合某指标与预警指标上下界间函数关系,从而量化预警指标的不确定性区间。进一步,提出了一种小概率事件计算方法,作为预警的依据。预警井位的地质和工程情况发生异常的可能性较高,可作为定期地质和工程测井的选井依据。

(3)建立的生产监测数据分析以及数据驱动的监测井网优化和预测预警方法对于促进智能油气田建设有重要意义,也有助于渗流机理不明确和地质模型不确定性高的实际油气藏的高效开发利用。

致谢:感谢中国石化胜利油田分公司油藏动态监测中心相关技术专家的大力支持,感谢孤岛项目部提供的现场试验帮助。

参考文献

- [1] Kikani J. Reservoir surveillance[M]. Society of Petroleum Engineers, 2013.
- [2] 刘爱侠. 海上平台油气生产预警方法研究[J]. 中国石油和化工标准与质量, 2017, 37(3): 24-25.
LIU Aixia. Research on early warning methods for oil and gas production on offshore platforms[J]. China Petroleum and Chemical Standard and Quality, 2017, 37(3): 24-25.
- [3] 赵智勇, 刘志斌, 徐建平. 油田监测指标预测预警开发动态系统[M]. 石油工业出版社, 2010: 58-68.
- [4] Alessa S. Production allocation of reservoir layers using data-driven production allocation of reservoir layers using data-driven reservoir modeling[D], West Virginia University, Graduate Thesis, 2020.
- [5] 李航. 统计学习方法[M]. 清华大学出版社, 2012.
- [6] 李建丽, 钟仪华, 刘道杰, 等. 小波神经网络的油田开发动态指标预测方法[J]. 天然气勘探与开发, 2008, 31(3): 62-66.
LI Jianli, ZHONG Yihua, LIU Daojie, et al. A method for predicting dynamic indicators of oilfield development based on wavelet neural network[J]. Natural Gas Exploration and Development, 2008, 31(3): 62-66.
- [7] 马承杰. 基于循环神经网络的油藏产量预测方法[J]. 内蒙古石油化工, 2021, 47(5): 108-112.
MA Chengjie. A method for predicting reservoir production based on recurrent neural networks[J]. Inner Mongolia Petrochemical Industry, 2021, 47(5): 108-112.
- [8] 同晓. 基于数据挖掘的油田开发动态预警系统研究[D]. 西安石油大学, 2016.
TONG Xiao. Research on dynamic early warning system for oilfield development Based on data Mining [D]. Xi'an Shiyou University, 2016.
- [9] CAO Chong, JIA Pin, CHENG Linsong, et al. A review on application of data-driven models in hydrocarbon production forecast[J]. Journal of Petroleum Science and Engineering, 2022, 212: 110-296.
- [10] Esmaili S, Mohaghegh S D. Full field reservoir modeling of shale assets using advanced data-driven analytics[J]. Geoscience Frontiers, 2016, 7: 11-20.
- [11] Ghassemzadeh S, Perdomo MG. A data-driven reservoir simulation for natural gas reservoirs [J]. Neural

Computing and Applications, 2021, 33 (18): 11777-11798.

- [12] 徐瑞萍. 一种多参数组合预警系统在油田生产中的应用[J]. 承德石油高等专科学校学报, 2019, 21(6): 16-18.
XU Ruiping. Application of a multi-parameter combined early warning system in oilfield production[J]. Journal of Chengde Petroleum Junior College, 2019, 21(6): 16-18.
- [13] 张严, 郑亚峰, 孙四维, 等. 基于数据驱动的长庆油田致密砂岩压后产量预测[J]. 能源与环保, 2021, 43(10): 96-101.
ZHANG Yan, ZHENG Yafeng, SUN Siwei, et al. Data-driven forecast of dense sandstone in Changqing oilfield [J]. Energy and Environmental Protection, 2021, 43(10): 96-101.
- [14] 钟仪华, 王淑宁, 罗兰, 等. 用深度学习挖掘油田开发指标预测模型的知识[J]. 西南石油大学学报: 自然科学版, 2020, 42(6): 163-74.
ZHONG Yihua, WANG Shuning, LUO Lan, et al. Mining the knowledge of the oilfield development index prediction model with deep learning [J]. Journal of Southwest Petroleum University: Natural Science Edition, 2020, 42(6): 63-74.
- [15] Turner SWD, Doering K, Voisin N. Data-driven reservoir simulation in a large-scale hydrological and water resource model[J]. Water Resources Research, 2020, 56(10): 1-16.
- [16] 周志华. 机器学习[M]. 清华大学出版社, 2016: 25-35.
- [17] 白晓虎, 姜汉桥, 王硕亮, 等. 油田开发动态指标多步预测模型研究[J]. 断块油气田, 2010(3): 345-347.
BAI Xiaohu, JIANG Hanqiao, WANG Shuoliang, et al. Study on multi-step prediction model for dynamic index of oilfield development[J]. Fault-Block Oil and Gas Field, 2010(3): 345-347.
- [18] 马立平, 任宝生, 赵明. 油田水驱开发动态系统预警指标体系的构建[J]. 特种油气藏, 2010, 17(4): 62-65.
MA Liping, REN Baosheng, ZHAO Ming. Construction of early warning index system of dynamic system of oilfield water flooding development [J]. Special Oil & Gas Reservoirs, 2010, 17(4): 62-65.

编辑 方志慧

第一作者简介:李建荣,男,1970年出生,教授级高级工程师,硕士研究生,1992年毕业于长春地质学院勘察地球物理专业,长期从事油气田开发和油藏动态监测研究工作。电话:0546-8680777, Email: Lijianrong109. slyt@ sinopec. com。通信地址:山东省东营市东营区胜利油田分公司油藏动态监测中心,邮政编码:257001。