

附加权重的LSTM神经网络页岩气产量预测

——以HS井区为例

甘世泽¹, 谭鑫龙²

¹重庆科技大学石油与天然气工程学院, 重庆

²潮州深能燃气有限公司, 广东 潮州

收稿日期: 2025年6月12日; 录用日期: 2025年7月10日; 发布日期: 2025年7月24日

摘 要

现阶段机器学习作为页岩气产量预测的一种有效手段, 在数据规模受限的开发新区却往往无法进行准确预测。针对此问题, 提出了附加权重的LSTM神经网络预测应用。首先使用HS页岩气井区地质参数、工程参数和生产参数的训练数据建立标准LSTM预测模型, 获得产量预测拟合公式; 其次通过机器学习方法交叉分析确定影响因素对测试产量的影响权重, 然后在初始LSTM模型引入附加权重来优化预测矩阵并进行产量预测, 最后应用模型对区块老井与新井的各项参数进行优化。预测结果表明, 附加权重后的LSTM模型预测值与真实值吻合度更高, 优化结果可靠, 可以很好地解决数据量受限井区的机器学习分析与优化问题。

关键词

页岩气, 机器学习, 产量预测, LSTM, 附加权重

Shale Gas Production Forecasting Using Additional Weights LSTM Neural Network

—A Case Study of the HS Well Area

Shize Gan¹, Xinlong Tan²

¹School of Petroleum Engineering, Chongqing University of Science and Technology, Chongqing

²Chaozhou Shenneng Gas Co., Ltd., Chaozhou Guangdong

Received: Jun. 12th, 2025; accepted: Jul. 10th, 2025; published: Jul. 24th, 2025

Abstract

Machine learning serves as an effective means for shale gas production forecasting at the current

文章引用: 甘世泽, 谭鑫龙. 附加权重的 LSTM 神经网络页岩气产量预测[J]. 矿山工程, 2025, 13(4): 836-845.

DOI: 10.12677/me.2025.134094

stage. However, accurate predictions are often challenging in newly developed areas with limited data. To address this issue, an additional weights Long Short-Term Memory (LSTM) neural network prediction approach is proposed. Initially, a standard LSTM prediction model is established using a training dataset comprising geological parameters, engineering parameters, and production parameters from the HS shale gas well area, obtaining a production prediction fitting formula. Subsequently, a machine learning approach is employed for cross-analysis to determine the impact weights of influencing factors on test production. These weights are then introduced into the initial LSTM model as additional weights to optimize the prediction matrix and enhance production forecasting. The results indicate that the LSTM model with additional weights yields predictions that closely align with real values, yielding reliable optimization results and providing an effective solution to machine-learning analysis and optimization in data-scarce well blocks.

Keywords

Shale Gas, Machine Learning, Production Forecasting, LSTM, Additional Weights

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

页岩气的开发在近年来得到了广泛的关注和重视, 不仅可以为国家提供可靠的能源供应, 增加经济收益和就业机会, 还可以减少温室气体排放和环境污染, 提高国际能源地位和影响力, 从而对能源市场、环境和地缘政治等方面产生重大影响, 具有重要的战略意义[1] [2]。

在大数据分析的不断发展下, 通过机器学习的页岩气产量预测方法被越来越多的人使用, 如 2020 年陈娟等运用 GA-BP 神经网络方法对长宁地区页岩气井进行预测和 2023 年韩克宁等通过产量递减与 LSTM 耦合的方法对常压页岩气井进行了产量预测[3] [4]。这些研究所使用的方法需要大量数据的支持, 在数据量受限的新井区往往效果不佳。

针对区块数据量较少的情况, 本研究阐述了一种附加权重的 LSTM 神经网络优化模型预测页岩气产量的方法。通过初始的 LSTM 模型分析现场实际生产数据获得各项因素对产量的影响, 然后使用机器学习方法交叉分析影响因素对测试产量的影响权重, 并将其作为附加权重对初始 LSTM 模型的预测矩阵进行优化, 提高模型预测值与真实值的吻合度, 为现场页岩气开发与调整提供参考。

2. 附加权重的 LSTM 神经网络模型

在传统的 LSTM 模型中, 各输入特征在时间步的处理过程中具有相同的重要性, 难以有效区分关键特征与冗余信息, 从而影响模型的学习效率与记忆能力。为提升模型对输入信息的识别能力, 研究者引入了附加权重机制, 通过对输入特征施加差异化权重, 引导模型更加关注对任务有显著贡献的特征。在这种机制下, 原始输入向量通过与附加权重向量的逐元素相乘实现加权处理, 使得高权重特征在门控单元的计算中占据主导地位, 进而增强其对记忆单元状态的影响。此外, 若附加权重作用于时间步维度, 形式上类似于注意力机制, 能够在多个时间节点的隐藏状态中动态赋权, 从而突出关键时间步对最终输出的贡献。这种加权机制不仅提高了模型对重要信息的选择性记忆能力, 也增强了其对长期依赖关系的建模效果, 显著改善了模型的表达能力与预测精度。

LSTM 的核心是通过记忆单元和门控机制来分析时间序列关系。包括遗忘门、输入门和输出门, 这

些门控制着信息的出入, 允许 LSTM 有选择地记住或忘记之前的信息, 从而有效地获取数据规律[5][6]。附加权重的 LSTM 模型是在输出门中添加一个参数权重, 以此来达到优化各项参数对输出控制的程度, 因此附加权重的 LSTM 模型在模型训练时与原生 LSTM 模型最大区别是使用附加权重对产量预测进行优化。

2.1. 数据收集与预处理

在本次研究中, 从现场收集已投入井的地质、钻井、水力压裂和试产生产等数据。首先将所有井的数据通过标准化的方法进行数据统一缩放, 对于少量缺失值采用插值补充, 大量缺失值则去除该参数, 而针对水力压裂参数存在的参数种类不统一的问题, 通过因素特征的选择与构建进行处理后得到综合数据表(表 1), 其中测试产量用于附加权重的确定[7]-[10]。

Table 1. Comprehensive data table
表 1. 综合数据表

数据分类		参数	参数值
静态参数	地质数据	有机碳含量(%)	2.49~5.70
		孔隙度(%)	4.60~24.11
		脆性指数(%)	51.82~76.63
		含气量(m ³ /t)	3.47~5.50
	工程数据	I + II类钻遇长度(m)	1241.00~1713.00
		平均压裂段长(m)	49.80~62.58
		施工排量(m ³ /m)	14.41~17.89
		加砂强度(m ³ /t)	2.58~4.08
		施工压力(MPa)	89.71~101.35
		用液强度(m ³ /t)	23.75~40.10
动态参数	试产数据	测试产量(×10 ⁴ m ³ /d)	4.30~21.82
	生产数据	生产时间(h)	0.00~24.00
		套压(MPa)	1.00~49.00
		输压(MPa)	0.00~7.85

2.2. 模型建立

遗忘门, 控制记忆单元状态中哪些信息需要保留或遗忘:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

其中, f_t 是遗忘门的输出, W_f 是权重矩阵, $[h_{t-1}, x_t]$ 表示前一时间步的隐藏状态和当前时间步的输入, $\sigma()$ 是激活函数。

输入门, 控制记忆单元状态中哪些新信息需要添加:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

其中, i_t 是输入门的输出, $\tilde{C}_t$ 是待更新的候选单元状态。

更新记忆单元状态:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

其中, C_t 是记忆单元状态, C_{t-1} 是前一时间步的记忆单元状态。

输出门, 控制哪些信息从记忆单元状态传递到隐藏状态, 为了通过附加权重实现 LSTM 模型的优化, 在预测输出时加入附加权重:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t^a] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

$$x_t^a = x_t \cdot W_a \quad (7)$$

其中, o_t 是输出门的输出, h_t 是当前时间步的隐藏状态, W_a 是附加权重。

2.3. 模型训练

使用 Python 语言进行代码编写来构建附加权重的 LSTM 模型, 并将预测井以外的其他井的静态参数 (除去测试产量) 与动态参数导入模型进行训练, 得到 loss 值 (训练误差) 如图 1。从图 1 可以看出, 在不断的训练中, 模型的预测值与真实值的误差在不断地减低, 最终达到良好的值并趋于稳定, 表示建立的 LSTM 模型良好, 可以用于分析与应用。

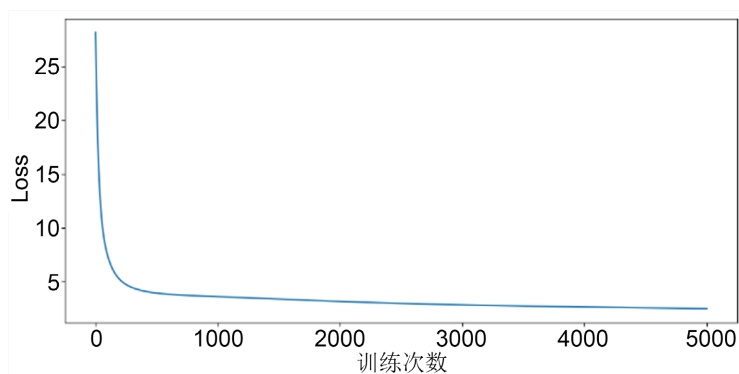


Figure 1. Loss curve of model training

图 1. 模型训练 loss 图

训练好模型后可以根据训练的特征与模型本身来计算各个输入特征的 SHAP 值 (特征对输出结果的贡献权重) 从而得到各个输入特征的权重如图 2 所示。

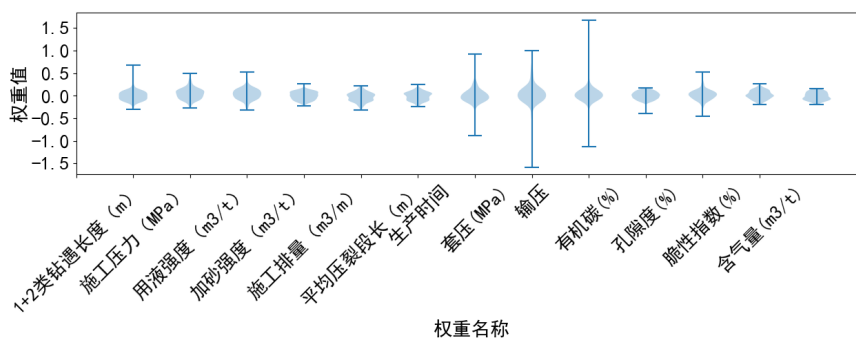


Figure 2. Distribution of LSTM model weights illustrated by violin plot

图 2. LSTM 模型权重值小提琴图

3. 附加权重的确定

本次研究中针对附加权重模型算法分别选择了集成学习的随机森林与非参数回归的支持向量机来比较分析, 以测试产量作为输出, 其他静态参数作为输入来训练模型, 并计算模型的 SHAP 值作为附加权重[11] [12]。

3.1. 权重分析模型

随机森林是一种集成学习方法, 它通过组合多个决策树来改善预测性能, 并且在机器学习中被广泛应用, 其预测结果可用以下公式表示[13] [14]:

对于分类问题, 采用投票法:

$$\hat{y}_{RF}(x) = mode(\hat{y}_1(x), \hat{y}_2(x), \dots, \hat{y}_n(x)) \tag{8}$$

对于回归问题, 采用平均法:

$$\hat{y}_{RF}(x) = 1/n \tag{9}$$

其中, $\hat{y}_{RF}(x)$ 是随机森林的预测结果, $\hat{y}_i(x)$ 是第 i 个决策树的预测结果, n 是决策树的数量。

支持向量机是一种用于二分类和多分类问题的算法, 其主要目标是找到一个能够将不同类别的数据分开的最优超平面, 其表达式可以表示为[15] [16]:

$$\omega \cdot x + b = 0 \tag{10}$$

其中, ω 是法向量(垂直于超平面的向量), x 是数据点的特征向量, b 是截距。对于给定的样本点, 可使用以下公式来进行分类:

$$f(x) = sign(\omega \cdot x + b) \tag{11}$$

这里, $f(x)$ 的符号表示数据点所属的类别, $sign$ 功能为判断 $(\omega \cdot x + b)$ 的结果, 如果大于 0, 则分为正类; 如果小于 0, 则被分为负类。

根据编译好的模型代码分别输入训练数据以及预测井的特征进行训练分析, 此处选择两口井的测试产量作为目标, 综合误差作为模型训练指标, 交叉验证得出最佳优化模型与模型的预测值和误差如表 2 所示。

综合误差计算公式:

$$e_{com} = \frac{(|q_{f1} - q_{t1}|) + (|q_{f2} - q_{t2}|)}{(q_{t1} + q_{t2})} \times 100\% \tag{12}$$

其中, e_{com} 是综合误差, q_f 是预测值, q_t 是真实值。

Table 2. Test yield prediction results of the best model
表 2. 最佳模型测试产量预测

井名	模型算法							
	随机森林				支持向量机			
	预测值 (10 ⁴ m ³)	真实值 (10 ⁴ m ³)	绝对 误差	综合误差 (%)	预测值 (10 ⁴ m ³)	真实值 (10 ⁴ m ³)	绝对 误差	综合误差 (%)
YH4-3	15.71	17.85	2.14	20.59	13.91	17.85	3.94	36.75
YH4-4	15.79	21.82	6.03		11.18	21.82	10.64	

3.2. 权重分析

同样通过计算最佳模型的 SHAP 值来获得各参数的权重分布(图 3), 并计算包含预测井特征最小范围的平均权重得到各参数的附加权重(表 3), 此附加权重可直接用于产量预测的优化应用。

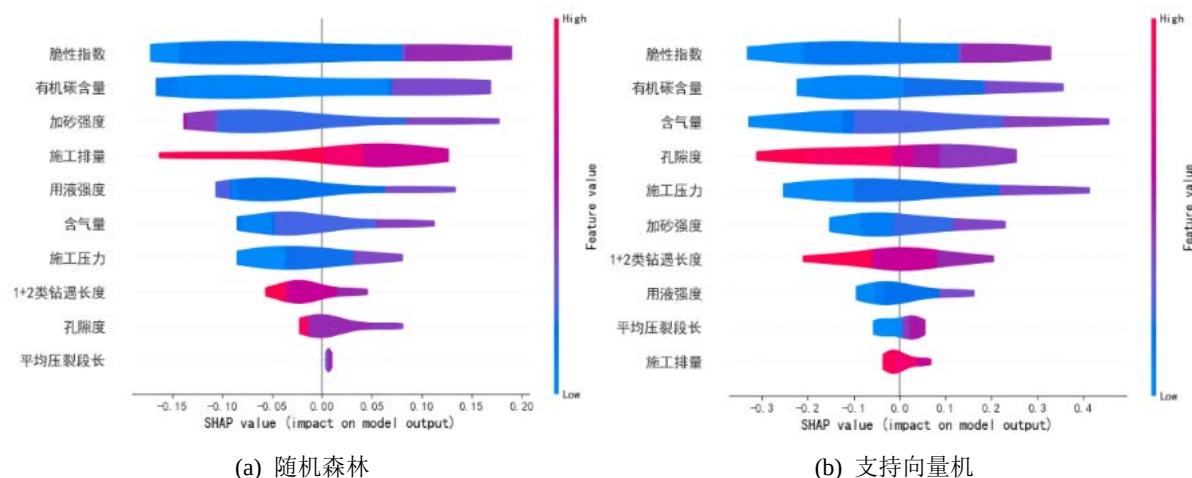


Figure 3. Distribution of weights

图 3. 权重分布图

Table 3. Additional weight table

表 3. 附加权重表

影响因素	平均权重	
	随机森林	支持向量机
有机碳含量	-0.015	0.096
孔隙度	0.011	0.172
脆性指数	-0.006	0.234
含气量	-0.013	0.016
I + II类钻遇长度	-0.014	0.001
平均压裂段长	-0.018	-0.159
施工排量	-0.027	-0.053
加砂强度	-0.034	-0.026
施工压力	0.022	-0.017
用液强度	0.006	0.022

4. 产量预测应用

4.1. 生产井产量预测

将各分析模型得到的特征权重作为附加权重输入 LSTM 验证模型的输入参数权重矩阵, 对两口井分别进行产量预测, 见图 4。

(1) YH4-3 井:

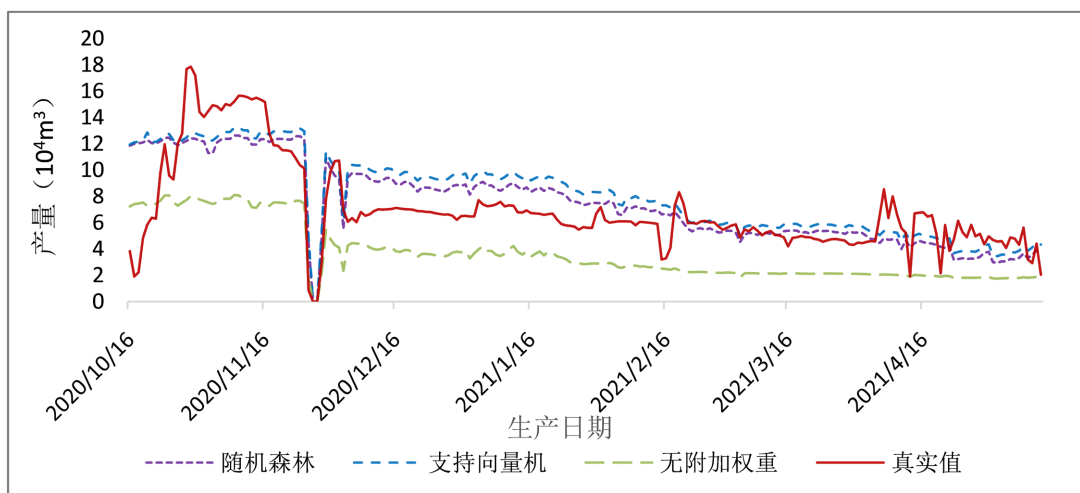


Figure 4. Predicted production for well YH4-3

图 4. YH4-3 井产量预测

根据各模型的预测产量与真实产量对比图分析可得 YH4-3 井附加随机森林模型的权重效果比附加支持向量机权重效果好, 见图 5。

(2) YH4-4:

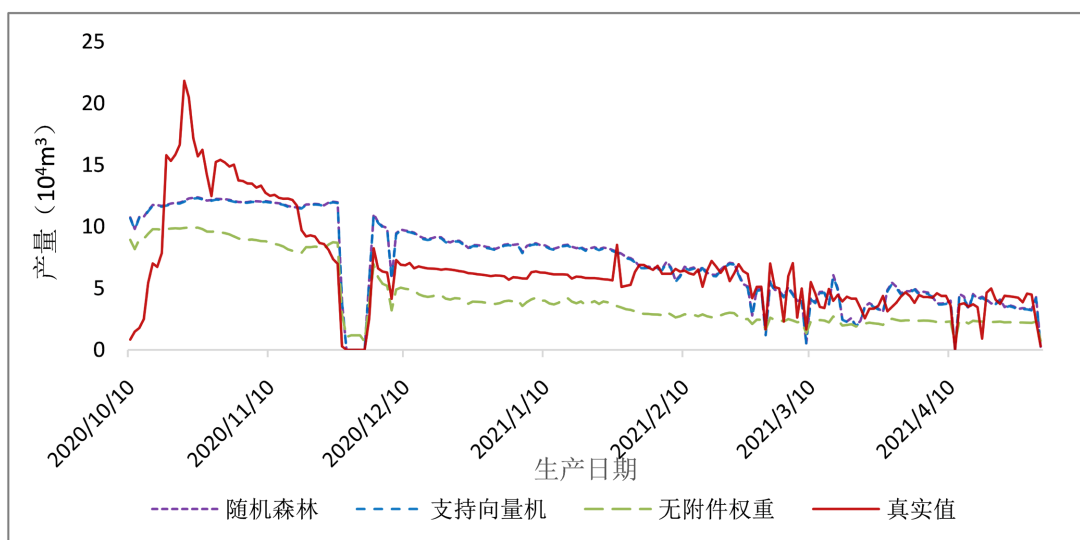


Figure 5. Predicted production for well YH4-4

图 5. YH4-4 井产量预测

根据各模型的预测产量与真实产量对比图可以看出 YH4-4 井附加随机森林权重和附加支持向量机权重效果相差不大。综合两口井的结果优选出随机森林模型的权重作为本井区的附加权重。

4.2. 老井生产参数优化

对于已经施工完毕并正进行生产中的老井, 其工程参数无法更改, 则需要对其生产参数进行优化, 也就是压力、油气渗流条件等参数进行优化, 例如生产时长与关井时长、输入补充压力的大小等。

以 YH4-3 井为例, 先进行输入补充压力大小的优化, 设定输入补充压力大小变化范围为 1~7 MPa,

生产时间 1 个月, 通过模型循环预测不同输压大小下 1 个月的平均产量与评价压降, 综合优选出最佳输压范围。将分组后的数据代入模型得到各值下的压降与产量如图 6。

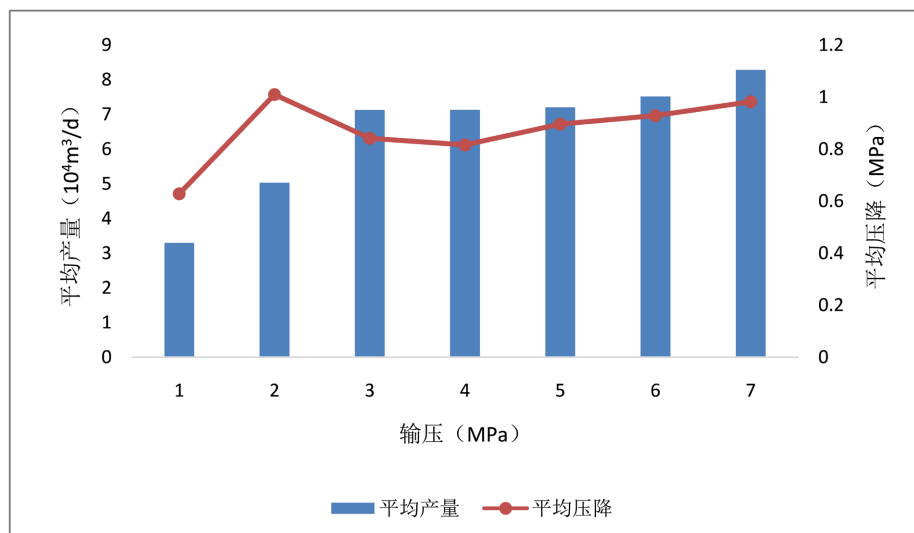


Figure 6. One-month average production of well YH4-3 under different delivery pressures

图 6. YH4-3 不同输压下的一个月平均产量

从图 6 中可以看出 YH4-3 井在输入压力为 3~5 MPa 时得到的平均产量较高, 平均压降较低, 因此其最佳输压范围为 3~5 MPa。从此范围中选取 4 MPa 的输压作为接下来的最佳生产时长与关井时长比例优化的输压。

设定生产时间 1 个月, 输压 4 MPa, 生产时长与关井时长比例从 1:1 到 5:1, 综合分析各比例的平均产量与平均压力变化优选出最佳生产时长与关井时长比例。将分组后的数据代入模型得到各比例下的压力变化与产量变化如图 7。

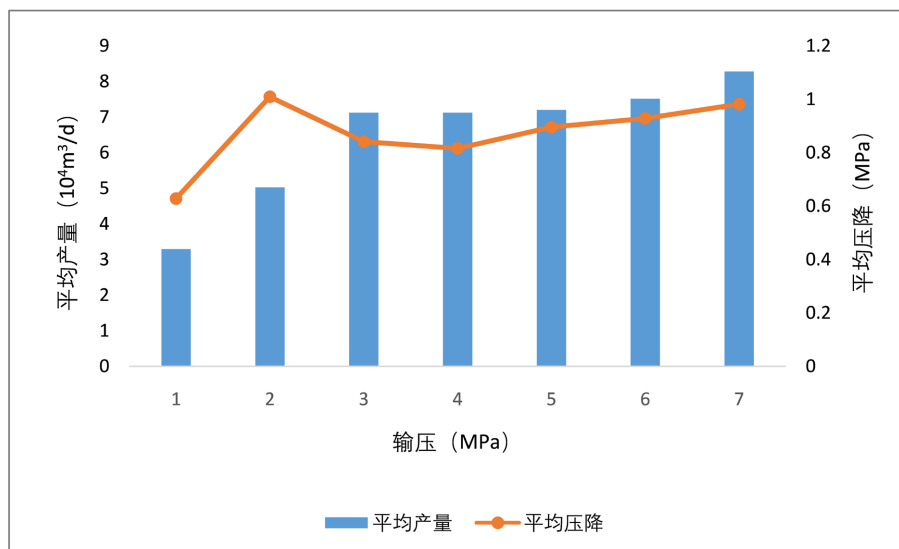


Figure 7. Average production and pressure variation of well YH4-3 under different ratios of production time to shut-in time

图 7. YH4-3 井各生产时长与关井时长比例下的平均产量与压力变化

从图 7 可以得到在生产时长与关井时长比例为 1:1 到 2:1 时压力恢复和平均产量较高, 但生产后的压降较大, 不适宜长期稳定生产; 在比例为 3:1 到 5:1 时压力恢复与压降都较为稳定, 再根据平均产量的大小优选出最佳生产时长与关井时长比例为 2:1 到 3:1 之间。

最终对于 YH4-3 井的生产参数优化结果为输压范围在 3~5 MPa, 生产时长与关井时长比例在 2:1 到 3:1。

4.3. 粒子群优化

粒子群优化(PSO)是一种优化算法, 基本思想是通过不断调整每个粒子的各项参数来寻找最优结果。对于已经训练好的预测模型, 粒子群各出入参数的更新受到粒子自身历史最优解和全局最优解的影响, 从而调整各项参数值, 以追寻到最优的输出结果。

针对新开发井, 选取 YH4-3、YH4-4 等井为参考井, 将 I + II 类储层钻遇长度设定为各参考井的参数数值, 并且选定平均压裂段长、加砂强度、施工压力、用液强度和施工排量等工程参数的参数变化范围, 通过粒子群优化的方法对分析模型的输入参数(上述工程参数)进行优化, 得到最高输出结果(生产动态产能)下的各项输入参数, 优化结果如表 4 所示。从表 4 中可以得到 1000 m 到 1500 m 之间不同的 I + II 类储层钻遇长度条件下生产井的各工程参数最优数值, 此优化结果可作为新生产井的工程参数设计与调整参考。

Table 4. Optimization results of engineering parameters for the reference well
表 4. 参考井工程参数优化结果

参数	I + II 类钻遇 长度 (m)	施工 压力 (MPa)	用液 强度 (m ³ /t)	加砂 强度 (m ³ /t)	施工 排量 (m ³ /m)	平均压裂 段长 (m)	测试产量 (10 ⁴ m ³ /d)	产能 增长 (%)
随机 范围		[80, 120]	[20, 50]	[2, 5]	[10, 30]	[30,90]		
YH 4-3	1469.00	90.98	30.04	3.16	17.62	60.00	9.09	
优化 结果	1469.00	119.40	30.91	3.40	27.13	46.20	16.84	85.25
YH 4-4	1241.00	94.39	33.25	4.08	17.70	62.58	8.98	
优化 结果	1241.00	109.24	49.56	4.77	22.07	82.67	16.75	86.52

5. 结论

使用 LSTM 模型虽然能够学习到页岩气产量递减的规律, 但在井数据量较少的区块, 单纯的 LSTM 模型并不能很好适应, 而附加权重后能提高模型的适应性, 大大降低预测的误差。在页岩气开发现场可以根据需求调整输入的特征条件, 从而得到不同参数下的预测产量, 并通过获得更高预测产量时的参数配置来优化老井与新井的参数, 为页岩气的开发提供参考。

参考文献

[1] 邹才能, 董大忠, 王社教, 等. 中国页岩气形成机理、地质特征及资源潜力[J]. 石油勘探与开发, 2010, 37(6): 641-653.
[2] 张金川, 徐波, 聂海宽, 等. 中国页岩气资源勘探潜力[J]. 天然气工业, 2008(6): 136-140+159-160.

- [3] 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): 1851-1858.
- [4] 韩克宁, 王伟, 樊冬艳, 等. 基于产量递减与 LSTM 耦合的常压页岩气井产量预测[J]. 油气藏评价与开发, 2023, 13(5): 647-656.
- [5] Chen, X., Li, J., Gao, P. and Zhou, J. (2022) Prediction of Shale Gas Horizontal Wells Productivity after Volume Fracturing Using Machine Learning—An LSTM Approach. *Petroleum Science and Technology*, **40**, 1861-1877. <https://doi.org/10.1080/10916466.2022.2032739>
- [6] Kocoglu, Y., Gorell, S. and McElroy, P. (2021) Application of Bayesian Optimized Deep Bi-LSTM Neural Networks for Production Forecasting of Gas Wells in Unconventional Shale Gas Reservoirs. *Proceedings of the 9th Unconventional Resources Technology Conference*, Houston, 26-28 July 2021, 2176-2196. <https://doi.org/10.15530/urtec-2021-5418>
- [7] Wang, T., Wang, Q., Shi, J., Zhang, W., Ren, W., Wang, H., et al. (2021) Productivity Prediction of Fractured Horizontal Well in Shale Gas Reservoirs with Machine Learning Algorithms. *Applied Sciences*, **11**, Article 12064. <https://doi.org/10.3390/app112412064>
- [8] Yang, R., Liu, X., Yu, R., Hu, Z. and Duan, X. (2022) Long Short-Term Memory Suggests a Model for Predicting Shale Gas Production. *Applied Energy*, **322**, Article 119415. <https://doi.org/10.1016/j.apenergy.2022.119415>
- [9] Zhai, S., Geng, S., Li, C., Gong, Y., Jing, M. and Li, Y. (2022) Prediction of Gas Production Potential Based on Machine Learning in Shale Gas Field: A Case Study. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, **44**, 6581-6601. <https://doi.org/10.1080/15567036.2022.2100521>
- [10] Hui, G., Chen, S., He, Y., Wang, H. and Gu, F. (2021) Machine Learning-Based Production Forecast for Shale Gas in Unconventional Reservoirs via Integration of Geological and Operational Factors. *Journal of Natural Gas Science and Engineering*, **94**, Article 104045. <https://doi.org/10.1016/j.jngse.2021.104045>
- [11] Marcilio, W.E. and Eler, D.M. (2020) From Explanations to Feature Selection: Assessing SHAP Values as Feature Selection Mechanism. 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Porto de Galinhas, 7-10 November 2020, 340-347. <https://doi.org/10.1109/sibgrapi51738.2020.00053>
- [12] Van den Broeck, G., Lykov, A., Schleich, M. and Suciu, D. (2022) On the Tractability of SHAP Explanations. *Journal of Artificial Intelligence Research*, **74**, 851-886. <https://doi.org/10.1613/jair.1.13283>
- [13] 谢兆贤, 邹兴敏, 张文静. 大型数据集的高效参数剪枝决策树算法研究[J]. 计算机工程, 2024, 50(1): 156-165.
- [14] Safavian, S.R. and Landgrebe, D. (1991) A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, **21**, 660-674. <https://doi.org/10.1109/21.97458>
- [15] 李福祥, 王雪, 张驰, 等. 基于边界点的支持向量机分类算法[J]. 陕西理工大学学报(自然科学版), 2022, 38(3): 30-38.
- [16] Wang, H. and Hu, D. (2005) Comparison of SVM and LS-SVM for Regression. 2005 *International Conference on Neural Networks and Brain*, Beijing, 13-15 October 2005, 279-283. <https://doi.org/10.1109/icnnb.2005.1614615>