引用格式：胡义升，罗盛源，黄昊，等. 页岩油水平井产量智能模拟预测：以鄂尔多斯盆地三叠系延长组7段为例[J]. 油气藏评价与开发，2026，16(6)：1286-1296.
Hu Yisheng, Luo Shengyuan, Huang Hao, et al. Intelligent simulation and prediction of shale oil production in horizontal wells: a case study of Chang 7 shale oil of Triassic System, Ordos Basin[J]. Petroleum Reservoir Evaluation and Development, 2026, 16(6): 1286-1296.
DOI: 10.13809/j.cnki.cn32-1825/te.2025441

页岩油水平井产量智能模拟预测
——以鄂尔多斯盆地三叠系延长组7段为例

胡义升[1]，罗盛源[1]，黄 昊[2]，陈继宏[1]，黄天镜[3]，张嘉辉[4]
（1.西南石油大学石油与天然气工程学院，四川 成都 610500；
2.中国石油长庆油田分公司第五采油厂，陕西 西安 710299；3.中国石油长庆油田分公司勘探开发研究院，陕西 西安 710018；
4.西南石油大学计算机科学学院，四川 成都 610500）

摘要：鄂尔多斯盆地三叠系延长组7段（长7段）页岩油藏因其储层横向非均质性显著，原油物性变化明显，一直未实现规模化有效开发。准确预测单井产量并确定其主控因素对优化现场生产措施至关重要。基于长7段筛选出的197口样本井的生产数据、工程和压裂参数，首先利用特征加权模糊C均值聚类（Fuzzy C-Means Clustering，FCM）算法对单井进行分类，并构建引入注意力机制的长短期记忆网络（Long Short-Term Memory，LSTM）模型进行产量智能模拟预测。该机制能够有效克服LSTM模型在多参数条件下筛选关键信息的难题，优先考虑最为关键的要素，忽视影响较小的要素；再通过差分进化算法对模型进行参数优化，最后将产量预测值与LSTM模型和传统Arps模型预测值进行对比分析。结果表明：在经过分类后的3类单井中，含油饱和度、压裂段数、加砂量、入地总液量以及油层长度等是影响单井产量的主要因素，且每个因素所占比重在3类单井中各不相同。传统的Arps递减模型在决定系数（R^2）、均方根误差（RMSE）和平均绝对百分比误差（MAPE）这3个指标上均体现出较大的误差，在页岩油藏单井产量预测上没有神经网络模型预测效果好。在引入注意力机制后，LSTM模型在3类单井产量的预测精度上分别提升了3.34%、3.57%和2.57%，表明注意力机制良好的优化效果。根据3类单井在不同生产阶段的特点，调整AT-LSTM模型的输入参数和训练集长度能够进一步提高产量预测效果。以提高单井产量为目标，对区域内单井进行主控因素的优化，证实了多因素协同优化的可行性，同时为长7段页岩油藏单井的生产措施调整提供了有力支持。

关键词：鄂尔多斯盆地；延长组7段；页岩油；注意力机制；差分进化算法；LSTM模型；产量预测
中图分类号：TE319 **文献标识码：**A

Intelligent simulation and prediction of shale oil production in horizontal wells: a case study of Chang 7 shale oil of Triassic System, Ordos Basin

Hu Yisheng[1], Luo Shengyuan[1], Huang Hao[2], Chen Jihong[1], Huang Tianjing[3], Zhang Jiahui[4]

*(1. School of Petroleum and Natural Gas Engineering, Southwest Petroleum University, Chengdu, Sichuan 610500, China;
2. No.5 Oil Production Plant of PetroChina Changqing Oilfield Company, Xi'an, Shaanxi 710299, China;
3. Exploration and Development Research Institute of PetroChina Changqing Oilfield Company, Xi'an, Shaanxi 710018, China;
4. School of Computer Science, Southwest Petroleum University, Chengdu, Sichuan 610500, China)*

Abstract: Chang 7 shale oil of the Triassic System in the Ordos Basin have not yet achieved large-scale and effective development due to their significant lateral reservoir heterogeneity and considerable variations in crude oil physical properties. Accurately predicting single-well production and identifying its main controlling factors are crucial for optimizing field production strategies. Based on production data, engineering parameters, and fracturing parameters from 197 sample wells screened in the Chang 7 shale oil, this study first classified the wells using a feature-weighted fuzzy C-means clustering (FCM) algorithm. Subsequently, a long short-term memory (LSTM) model incorporating an attention mechanism was constructed for intelligent production simulation and prediction. This mechanism effectively overcomes the difficulty of the standard LSTM model in screening key information under multi-parameter conditions by prioritizing the most critical factors and ignoring less influential ones. The model parameters were then optimized using a differential evolution algorithm. Finally, the production predictions were compared against those from the standard LSTM model and the traditional Arps model. The results indicated that across the three classified types of single wells, the main factors influencing single-well production were oil saturation, the number of fracturing stages, sand volume, the total injected fracturing fluid volume, and reservoir length. Furthermore, the relative contribution of each

收稿日期：2025-09-16。
第一作者简介：胡义升（1988—），男，博士，研究员，主要从事注气提高采收率、CO_2埋存、油气藏流体相平衡及高压物性实验等研究。地址：四川省成都市新都区新都大道8号，邮政编码：610500。E-mail：yisheng.hu@swpu.edu.cn
基金项目：四川省重大专项“基于大数据的稠油开发动态模拟预测及优化调控技术”（20QYCX0030）。

factor varied among the three well types. The traditional Arps decline model showed relatively large errors in three evaluation indicators—coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE)—demonstrating that neural network models provided superior prediction performance for single-well production in this shale oil reservoir compared to the Arps model. After introducing the attention mechanism, the prediction accuracy of the LSTM model for the production of three well types improved by 3.34%, 3.57%, and 2.57%, respectively, demonstrating its effective optimization capability. Adjusting the input parameters and training set length of the AT-LSTM model according to the characteristics of different production stages for the three well types can further enhance production prediction performance. With the objective of increasing single-well production, optimization of the main controlling factors for single wells within the block confirmed the feasibility of multifactor synergistic optimization, while providing strong support for adjusting production strategies for single wells in the Chang 7 shale oil reservoir.

Keywords: Ordos Basin; Chang 7 member of Yanchang Formation; shale oil; attention mechanism; differential evolution algorithm; LSTM model; production prediction

页岩油作为21世纪新兴的一种非常规油气资源,其大规模开发是减少对传统石油资源依赖的重要手段[1-2]。中国的页岩油以陆相为主,已探明资源量约1 500×10^8 t,可采资源量约40×10^8 t,具有巨大的资源潜力[3-4]。鄂尔多斯盆地三叠系延长组7段(简称长7段)页岩油目前勘探开发的主要对象是夹层型页岩油,受陆相湖盆气候条件、构造条件复杂等因素影响,整体呈现非均质性强、原油性质变化大等特征,规模效益开发面临极大挑战[5]。在考虑页岩油单井的地质、工程以及生产动态参数的前提下,如何利用现场生产数据对单井产量进行有效预测,是长7段油层能否进行大规模开发的基础。

单井产量预测方法主要分为曲线递减法、数值模拟方法、定量表征法和机器学习方法。陈劲松等[6-8]提出了一种基于Arps递减模型的改进模型,对4种工程中常用递减预测模型(Arps超双曲递减、Arps调和递减、扩展指数递减和DUONG递减)在不同生产阶段的预测进行了详细的对比分析。董鹏等[9]建立了多裂缝直井产能模型,运用保角变换和Pedrosa替代法求解解析产能模型,该方法考虑了多种外部条件,因此,被广泛运用于计算压裂水平井单相流体产能,但无法准确描述实际油藏的生产状况。杨洋等[10]通过时间序列分析结合残差修正方法,建立具有动态分析能力的产量差分自动回归移动平均模型,得到了较LSTM模型而言更为准确的预测值。刘阳等[11]提出了一种基于多维时序长短期记忆神经网络的机械钻速预测方法,通过优选工程参数和其他参数的组合,确定了最优超参数组合方案,较好提升了预测效果。闵超等[12]提出一种基于FCMFS特征选择算法的煤层气压裂效果预测方法,通过遗传编程和XGBoost算法进行影响因素特征构造和筛选,能够提高模型预测准确率和召回率。

页岩油藏复杂的井下条件以及多因素的影响,使得常规方法难以准确预测不同条件下的页岩油产能。随着人工智能的迅猛发展,结合机器学习方法预测页岩油产能成为近年热门研究方向。官文洁等[13]通过灰狼算法优化长短期记忆网络(Long Short-Term Memory, LSTM)模型,对页岩油开采前期与后期产能分别进行模拟预测,该方法能够弥补LSTM神经网络模型参数多以及调优困难等问题,但只考虑实际生产数据参数对产能的影响,并未考虑工程参数以及地质参数的影响。

本研究通过特征加权模糊C均值聚类(Fuzzy C-Means Clustering, FCM)算法对长7段进行单井主控因素的分析,将区域内单井分为3类,主控因素主要包括含油饱和度、压裂段数、入地总液量、加砂量、油层长度、渗透率和压裂簇数。在此基础上利用Attention注意力机制使得LSTM模型能够模拟人类行为自动选取关键信息,再通过具有结构简单、收敛速度快特点的差分进化算法快速搜索出全局最优解。由此得到的智能模拟产量预测模型能够在长7段基础数据完善的条件下,考虑多种人为可控因素,快速预测出单井产量,为后续长7段页岩油的开发具有一定的指导意义。

1 长7段页岩油开发历程

庆城油田位于鄂尔多斯盆地西南部,地处甘肃省华池县、庆城县、合水县、宁县境内,北至车道,南至太平,西至三岔,东至桐川,面积约5 000 km^2。主力开发区为华池与合水,主力开发层位为三叠系延长组长7段页岩油层(图1)。长7段页岩油经过勘探发现、评价探索、开发试验、规模开发这4个阶段,主体技术和管理模式逐渐成熟[5]。

随着长7段的开发和研究深入,诸多问题暴露出来。储层物性差,非均质性强;脆性矿物含量高,压力系统异常;天然能量弱;埋藏深,区域差异性大等特征使得开发难度巨大。人工智能近年来的迅猛发展,推动了油气开发工程技术的革新。立足于长7段完整的基础数据优势,可以构建出优质的智能模拟模型,为该区域的高效开发提供有效支撑。

2 基于AT-LSTM神经网络的单井产量预测模型

2.1 长7段单井分类以及数据处理

2.1.1 数据预处理

通过对长7段页岩油藏现场数据的收集与整理,得

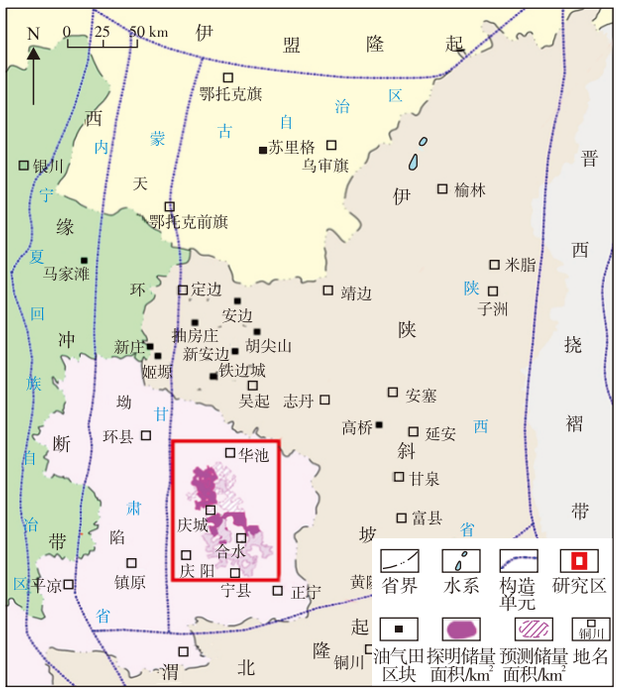


图1　鄂尔多斯盆地区域构造单元及庆城油田地理位置

Fig. 1　Regional tectonic units of Ordos Basin and geographical location of Qingcheng oilfield

到322口单井的原始数据。实际生产过程中，单井受到地质环境，开井和关井等操作的影响，原始生产数据会存在异常，不能直接用于建模。因此，需要对原始数据进行处理，具体步骤如下：

①部分单井5项地质参数均无数据，去除该情况下的84口单井；

②根据长7段所有页岩油井数据分析，当单井含水率高于60%时，单井的生产动态和产量衰减规律与其余产油井存在本质差异，若将其纳入会严重污染训练样本，因此，去除单井含水率高于60%的19口单井，至此已筛选出219口单井；

③部分单井地面原油黏度没有数据，将该单井所在平台的黏度值视为该单井的地面原油黏度；

④部分单井的平均全烃体积分数、总油层长度、孔隙度、含油饱和度、脆性指数、渗透率等数据缺失，由于缺项较少，采用简单平均的方法，分别取其参数现有数据的平均值补全。

为了进一步提升分类方法和建模的可靠性，对初筛的数据进行二次筛选，具体筛选步骤如下：

①去掉投产月数小于6个月的单井，这部分井生产时间太短，初期压裂液返排会导致其月产油量异常，影响总体样本质量；投产月数大于或等于6个月的单井，已经度过初期压裂液返排造成影响的阶段，其单井月产量能够纳入样本集中。

②去掉生产过程中（不包含未投产和关井后）超过50%的月产油量数据均为0的单井。

经上述筛选后，长7段页岩油藏共有197口单井可作为建模样本井，此举会导致后续产量预测模型无法完整地对长7段所有井做出有效且精准的单井产量预测，但会大幅提升可预测的单井产量的预测精度。

2.1.2　基于特征加权FCM算法的单井分类

传统FCM算法是一种采用目标函数进行模糊聚类的分析方法[14]，适用于数据聚类分析，核心思想是最大化同一类别内对象间的相似性，同时最小化不同类别间的相似性。由于传统FCM算法假设所有特征对聚类的贡献是相同的，这在实际中往往不成立。因此，特征权重系数$w=\{w_1,w_2,w_3,\cdots,w_n\}$通常被引入用来提高聚类效果，即特征加权FCM算法（Feature-Weighted Fuzzy C-Means Algorithm）[15]。在聚类过程中，不是所有特征都同等重要，通过为每个特征自动计算一个权重，来量化该特征对形成聚类结构的“贡献度”，从而得到更准确的聚类结果，进一步区分影响程度大小不同的特征维度。

$$J=\sum_{i=1}^{n}\sum_{k=1}^{c}u_{ik}^{m}\sum_{j=1}^{d}w_j^{\beta}\|x_{ij}-c_{kj}\|^2+\lambda\left(\sum_{j=1}^{d}w_j-1\right)\quad(1)$$

式中：J为目标函数；n为样本总数；c为聚类数目；u_{ik}为隶属度矩阵U的元素，表示第i个样本到第k个聚类中心的隶属度；m为隶属度矩阵U的加权指数，通常取值为2，表示隶属度的“模糊程度”；d为特征维度总数；w_j为第j个特征的权重；$\|x_{ij}-c_{kj}\|$为引入权重的欧几里得距离；β为控制权重分布，取值大于1；λ为拉格朗日乘子。

基于长7段筛选出的197口样本井各井的单井产量、含油饱和度、压裂段数、入地总液量和加砂量5个参数,其中单井产量是直接影响预测产量的因素;压裂段数、入地总液量和加砂量是工程参数中对产量影响最大的因素;地质参数则是含油饱和度。基于上述特征加权FCM算法对已选择的5个特征参数进行分类并获得对应的特征权重,如表1所示,后续表中产量均为单井第一年月均产油量。

通过上述几个特征权重值对长7段的单井进行综合加权评价,再根据欧几里得距离聚类分类后可知,样本井加权评分值聚类中心为3类时,单井聚类分类效果最好。因此,将197口样本井划分为Ⅰ、Ⅱ和Ⅲ类[16],部分类别单井数据如表2所示。

2.1.3 数据无量纲化处理

为了将影响单井产量的不同影响因素进行统一分析,此处通过最大-最小值(Max-Min)归一化方法将工程参数、地质参数和生产数据参数转化为同一标准,保留其原始数据的分布和形态特点,其数学表达式如式(2):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

式中:x'为归一化的数据,其值范围为[0,1];x为原始数据;$x_{\min}$为原始数据的最小值;$x_{\max}$为原始数据的最大值。

因此,对长7段已进行分类的单井通过归一化后的部分数据如表3—表5所示。

通过归一化数据结合现场生产措施分析可知:Ⅰ类井在长7段中具有最高的累积产油量、地层岩石孔渗最大,物性最好,现场采用了强有力的增产手段,因此,体现在表3中各项参数的值最高;Ⅱ类井物性中等,需要通过加大压裂措施的投入来弥补物性上的差异以保证获得可观的单井产量,因此,表4中工程参数值与Ⅰ类井接近;Ⅲ类井本身所处的地层条件较差,投入的成本也少,属于边际效益井,因此,在表5中工程和地质参数的值最低。现场开采主要

表1 特征加权FCM算法的不同参数特征权重结果

Table 1 Feature-weighted results for different parameters in weighted FCM algorithm

特征参数	产量	含油饱和度	压裂段数	入地总液量	加砂量
特征权重	0.525	0.214	0.103	0.093	0.065

表2 部分单井的特征加权评分

Table 2 Feature-weighted scoring for selected single wells

井号	类别	产量	含油饱和度	压裂段数	入地总液量	加砂量	加权评分值
华H16-2井	Ⅰ类	0.340	0.157	0.065	0.068	0.046	0.676
华H9-1井	Ⅰ类	0.375	0.149	0.027	0.029	0.022	0.603
华H40-4井	Ⅱ类	0.139	0.140	0.038	0.045	0.030	0.392
华H32-1井	Ⅱ类	0.373	0.022	0.033	0.044	0.031	0.503
华H17-4井	Ⅲ类	0.031	0.000	0.011	0.015	0.016	0.073
华H28-4井	Ⅲ类	0.156	0.024	0.046	0.045	0.032	0.303

表3 长7段页岩油藏Ⅰ类单井产量及影响因素的归一化数据

Table 3 Normalized data for production and influencing factors of class Ⅰ single wells in Chang 7 shale oil reservoir

井号	产量	油层长度	水平段长度	平均全烃体积分数	入地总液量	加砂量	压裂段数
阳平8井	0.84	0.24	0.40	0.25	0.08	0.10	0.14
西平238-75井	0.72	0.50	0.63	0.25	0.21	0.09	0.07

表4 长7段页岩油藏Ⅱ类单井产量及影响因素的归一化数据

Table 4 Normalized data for production and influencing factors of class Ⅱ single wells in Chang 7 shale oil reservoir

井号	产量	油层长度	水平段长度	平均全烃体积分数	入地总液量	加砂量	压裂段数
阳平10井	0.24	0.50	0.40	0.30	0.18	0.00	0.63
华H61-4井	0.65	0.43	0.28	0.24	0.85	0.49	0.31

表5 长7段页岩油藏Ⅲ类单井产量及影响因素的归一化数据

Table 5 Normalized data for production and influencing factors of class Ⅲ single wells in Chang 7 shale oil reservoir

井号	产量	油层长度	水平段长度	平均全烃体积分数	入地总液量	加砂量	压裂段数
华H45-1井	0.78	0.64	0.50	0.10	0.16	0.49	0.33
华H62-3井	0.72	0.64	0.21	0.18	0.28	0.13	0.25

以经济效益为目标，主要关注对象为Ⅰ、Ⅱ类井，单井产量预测则将能够产生基本效益的目标井纳入考量，因此，将Ⅲ类井生产数据也纳入样本数据集。

2.1.4　单井产量影响因素皮尔逊相关性分析

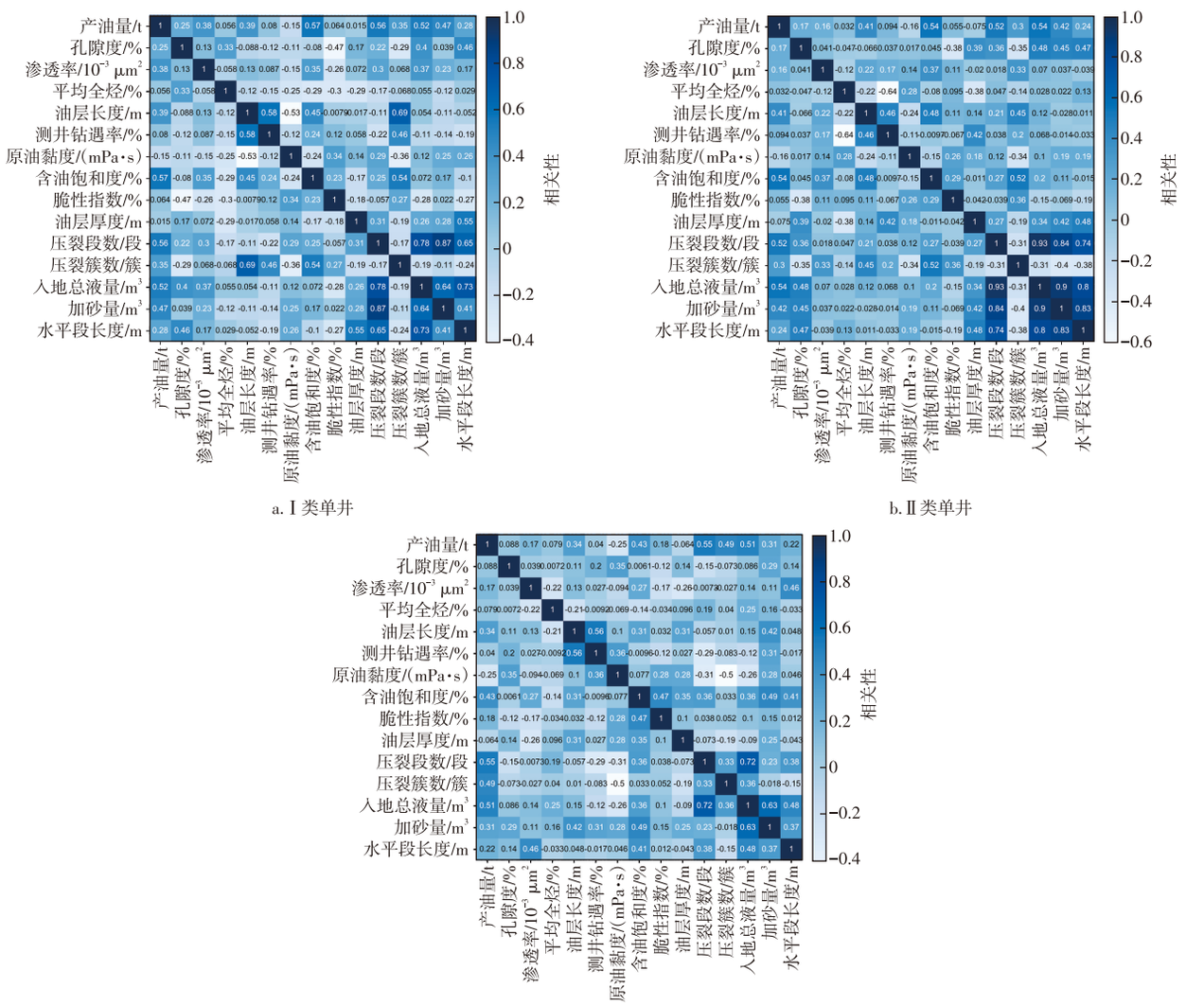
皮尔逊相关系数由Karl Pearson于19世纪末提出，常被用来彰显2个变量之间的线性关联度，其数学表达式如式(3)：

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{3}$$

式中：r为皮尔逊相关系数；x_i为比较数列的数据值；y_i为参考数列的数据值；$\bar{x}$为比较数列的均值；$\bar{y}$为参考数列的均值；n为样本数量。

影响单井产量的因素大都随开采时间的变化而改变，符合连续型变量的要求，通过对影响单井产量的不同因素进行统计分析，而后将归一化后的因素代入皮尔逊相关性分析的模型中[17]。图2为3类单井的皮尔逊相关性热力图，数值为正表示2个因素正相关，反之为负相关。通过分析图2中不同因素之间的关联性可知，r的绝对值接近1的因素为主要控制其分类井的关键人为可控因素。

其中，3类单井均含有含油饱和度、压裂段数、加砂量、入地总液量以及油层长度这几个因素，其平均相关性大小分别为0.51、0.54、0.40、0.52和0.38。通过对以上几个因素进行人为调控和干预，可以有效提高单井产量，为后续模拟时的可调整参数提供参考，除此之外的其他因素可以作为备选。如何在多种因素影响的条件下，有目的性地查找并自动调整其参数值成为建立产量预测模型的关键。



a. Ⅰ类单井　b. Ⅱ类单井　c. Ⅲ类单井

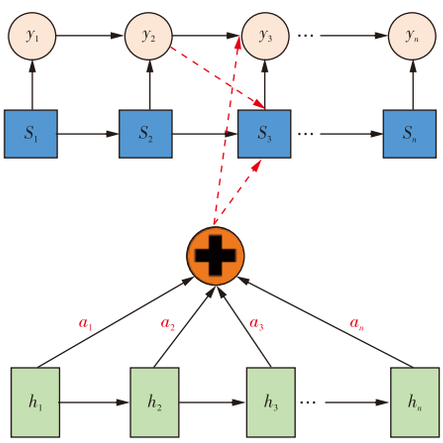
图2　Ⅰ—Ⅲ类单井产量与影响因素的关联关系热力图

Fig. 2　Heatmaps of correlations between production and influencing factors for class Ⅰ-Ⅲ single wells

2.2　AT-LSTM神经网络模型建立

2.2.1　Attention注意力机制

Attention注意力机制主要通过模仿人类的注意力行为，有针对性地关注输入数据中最相关的部分[18]。其核心思想是结合神经网络模型，使得模型能够增强特征提取能力，主要流程如图3。



注：⊕指注意力机制模型输入和输出的衔接部分，指代整个模型的名称；$\{h_1, h_2, h_3, \cdots, h_n\}$为输入序列的隐藏状态的特征值；$\{a_1, a_2, a_3, \cdots, a_n\}$为输入特征值对应的隐藏层的状态值；$\{S_1, S_2, S_3, \cdots, S_n\}$为最后输出时的隐藏层状态值；$\{y_1, y_2, y_3, \cdots, y_n\}$为经过注意力机制处理后得到的输出数据。

图3　Attention机制模型构架

Fig. 3　Architecture of attention mechanism model

注意力机制能够将多种因素同时输入并进行加权，有针对性地动态关注既定主控因素，为后续确定何时输出提供依据和支撑。其主要有2个算法步骤，第一步是对输入数据的权重分布进行计算，第二步则是进行加权平均值的计算，主要公式如式(4)—式(6)：

$$e_t = \tan h(wh_t + b) \tag{4}$$

$$a_t = \frac{\exp(e_t)}{\sum_{i=1}^{n} \exp(e_t)} \tag{5}$$

$$S_t = \sum_{i=1}^{n} e_t a_t \tag{6}$$

式中：e_t为t时刻输入数据的注意力概率分布值；h为隐藏状态映射向量；h_t为t时刻输入序列的隐藏状态特征向量；w与b分别为隐藏层的权重与偏置系数；a_t为历史输入的隐藏层状态对应的当前输入的权重值；S_t为隐藏层最终的输出值。

2.2.2　差分进化算法

差分进化(Differential Evolution，DE)算法，由Rainer Storn和Kenneth Price于1995年提出，于1997年正式发表期刊论文[19]，因其结构简单且能够快速收敛的特点常被用作一种自适应的全局优化算法。DE算法现被广泛用于人工神经网络、数据挖掘以及滤波器领域，其主要使用步骤包括种群初始化、变异操作、交叉操作及选择操作，最后迭代更新种群直到满足终止循环条件，迭代终止条件，即种群在连续多代中无明显改进时停止迭代，设定种群大小(N_P)、缩放因子(F)、基因维度(D)、交叉算子(C_R)和基因的取值范围[x_{min}，x_{max}]，具体过程如下。

1)种群初始化

在解空间内随机生成初始种群，其数学表达式如式(7)：

$$x_{i,j}^0 = x_{\min,j} + \text{rand}(0,1)\cdot(x_{\max,j} - x_{\min,j}) \tag{7}$$

式中：$x_{i,j}^0$为第0代第i个个体在第j维的初值($i = 1, 2, 3, ..., N_P$；$j = 1, 2, 3, ..., D$)；$x_{\min,j}$为第j维的下边界；rand(0,1)为均匀分布在[0,1]的随机数；$x_{\max,j}$为第j维的上边界。

2)变异操作

初始化后，对每一个目标种群个体进行变异操作生成相应的新变异个体，其数学表达式如式(8)：

$$v_i^g = x_{r1}^g + F\cdot(x_{r2}^g - x_{r3}^g) \tag{8}$$

式中：g为进化的迭代次数；v_i^g为第g代中第i个个体经变异后产生的变异个体；x_{r1}^g、x_{r2}^g、x_{r3}^g分别为第g代中编号为$r1$、$r2$、$r3$的基向量个体；F通常取值为[0,2]。

3)交叉操作

为了增加种群的多样性，需要用二项式交叉公式对变异产生的新个体进行交叉操作，其数学表达式如式(9)：

$$u_{i,j}^g = \begin{cases} v_{i,j}^g, \text{rand}(0,1) \leqslant C_R \text{或} j = j_{\text{rand}} \\ x_{i,j}^g, \text{其他情况} \end{cases} \tag{9}$$

式中：$u_{i,j}^g$为第g代中第i个个体在第j维上的试验个体；$v_{i,j}^g$为第g代中第i个个体在第j维上的变异个体；C_R通常取值为[0,1]；j_{rand}为随机选择的维度索引，以确保至少一个维度来自变异向量；$x_{i,j}^g$为第g代中第i个个体在第j维上的目标个体。

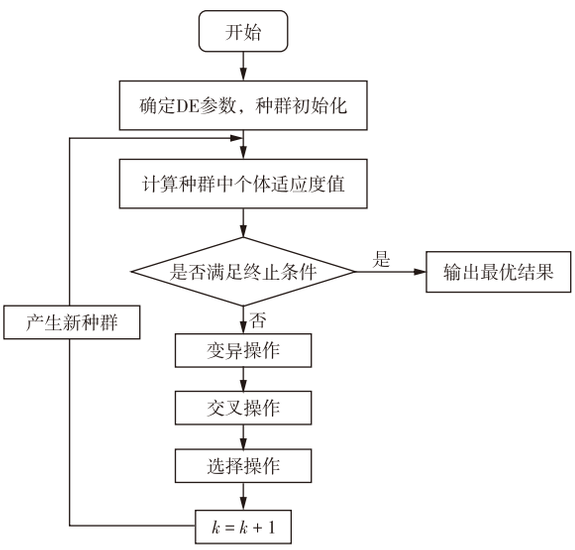
4)选择操作

为了对种群个体进行筛选和优化，需要将变异和交叉操作产生的试验个体与目标个体进行适应度函数对比，以选择更优个体，提高种群整体的适应度，避免过早收敛于局部最优解，其数学表达式如式(10)：

$$x_i^{g+1} = \begin{cases} u_i^g, f(u_i^g) \leqslant f(x_i^g) \\ x_i^g, \text{其他情况} \end{cases} \tag{10}$$

式中：x_i^{g+1}为第$g+1$代中第i个目标个体；$f(\cdot)$为适应度函数；u_i^g为第g代中第i个试验个体；x_i^g为第g代中第i个目标个体。

DE算法流程如图4所示。



注：k为算法的迭代次数。

图4　差分进化算法流程

Fig. 4　Flowchart of differential evolution algorithm

2.2.3　注意力机制的LSTM模型构建

近年来，随着人工智能的兴起，神经网络模型也逐渐被应用在油气开发领域。本研究针对常用LSTM模型无法处理长时间序列和多影响因素的数据集所暴露出来的局限性[20-21]，引入注意力机制来优化LSTM模型，使其有针对性地通过自主学习将注意力放在关键的影响因素上。为了提升AT(Attention-based)-LSTM模型性能并避免盲目地进行人工调参，采用DE算法对模型的超参数组合进行全局优化。算法主要优化目标是寻找最优的超参数(主要包括LSTM单元数、初始学习率及注意力机制中的权重系数等)，其适应度函数为均方误差(Mean Squared Error，MSE)，即通过最小化验证集上的MSE，使模型预测输出与实际产量的差异最小化，最终提升模型的预测精度与泛化能力。

本研究拟选用筛选出的197口样本井的单井产量数据，按照每口单井已生产时间的顺序排列，将单井前80%生产时间的产量作为训练数据，后20%生产时间的产量用于验证结果。在模型参数的具体配置中，将一次训练迭代中使用的样本数量(batch_size)设置为32，网格训练次数设置为1 000，LSTM单元数设置为16，初始学习率设置为0.000 5。其中DE算法关键参数的取值与范围设置，主要依据差分进化算法创始人Storn和Price的奠基性研究以及后续广泛的工程应用实践经验[19,22]。一般而言，N_p范围为单井产量预测维度的5～10倍，即在[50，200]的范围内进行测试，这样有助于在搜索效率与全局探索能力之间取得平衡；F作为控制差分变异的步长，区间设置为[0.4，1.0]，首先选择0.7，这样能有效避免在优化前期陷入局部最优，同时保持较好的收敛性；C_R设置在[0.7，1.0]这一范围，首先选择0.9，这样有助于将优良变异个体的信息引入种群，加速收敛至全局最优解。

模型的主要运行步骤如下，流程见图5：

①将样本数据输入模型前，对数据样本进行异常值剔除、插值填充和归一化处理，将不同因素统一计量；

②输入分类后单井的历史产量和各个影响因素，如含油饱和度、压裂段数、加砂量、入地总液量以及油层长度等；

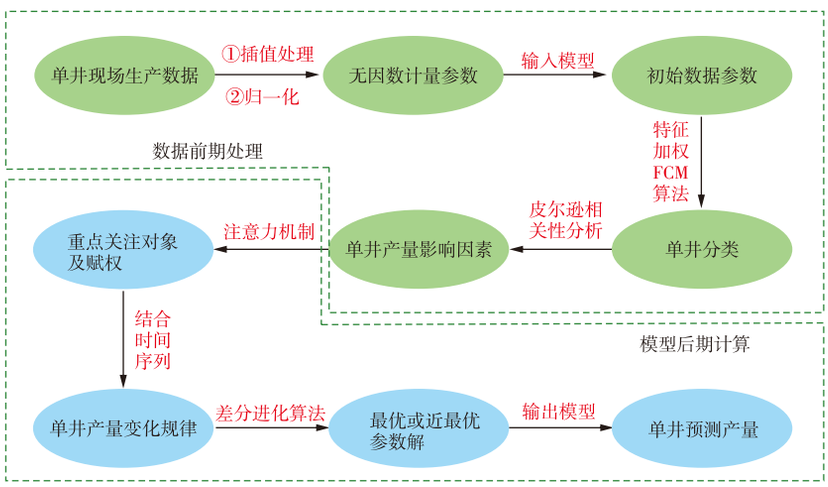


图5　基于注意力机制的LSTM模型预测产量流程

Fig. 5　Flowchart of attention mechanism-based LSTM model for production prediction

③通过注意力机制对输入的影响因素进行特征提取,自主选取重要的因素,并参照皮尔逊相关性系数表进行权重分配;

④通过LSTM神经网络结构的自适应学习功能,将影响因素和产量时间序列结合进行统计分析,得出规律;

⑤由LSTM产生的超参数通过DE算法找到最优参数解或者近乎最优参数解,并通过DE算法调节学习率和确定最佳的神经元数量;

⑥去掉部分过于理想化的数据,防止模型过拟合,提高模型的泛化性;

⑦将经由AT-LSTM模型处理后的数据输出为单井智能模拟预测产量。

3 实例应用及分析

3.1 模型评价标准

除了相较于原始生产数据的拟合精度(真实生产数据与模拟预测数据的比值)外,研究还拟采用决定系数(R^2)、均方根误差(Root Mean Squared Error,RMSE)和平均绝对百分比误差(Mean Absolute Percentage Error,MAPE)这3个指标来对模型的合理性作出评价和分析。其中,R^2是用来衡量拟合效果的直接方式,用于判断预测数据是否合理;RMSE可以反映真实数据与模型预测数据差异的间接体现;MAPE则能够评估模型的整体性能。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Q_{i,e} - Q_{i,p})^2}{\sum_{i=1}^{n}(Q_{i,e} - \bar{Q}_{i,e})^2} \tag{11}$$

$$\sigma_{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{i,e} - Q_{i,p})^2} \tag{12}$$

$$\sigma_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Q_{i,e} - Q_{i,p}}{Q_{i,e}}\right| \times 100\% \tag{13}$$

式中:$Q_{i,e}$为实际产量,单位t;$Q_{i,p}$为预测产量,单位t;$\bar{Q}_{i,e}$为平均实际产量,单位t;σ_{RMSE}为均方根误差,单位t;σ_{MAPE}为平均绝对百分比误差。

3.2 预测结果及分析

为了反映模型预测的精确性和适用性,此处引入一个常用的预测产量模型——Arps模型[7]。其作为经典的产量递减预测模型,被广泛用于油气藏产量预测,仅需要单井历史生产数据,但也因此在非常规页岩油的产量预测中具有一定的局限性。下面将从不同模型之间预测效果的对比和模型预测数据与实际数据之间的差异这两个方面来讨论模型优劣,后续结果分析中产量均为单井月产油量。

3.2.1 不同模型预测精度对比

分别从3类单井中各取一口井作为代表,用于此次产量预测模型预测精度的分析。通过分析3类单井各不相同的主要影响因素对产量的影响,能够进一步验证模型的适用性。表6为3类单井代表井在不同模型下的预测指标对比。

首先直观地观察R^2的大小,R^2值越接近1,模型预测效果越好。3类代表井中AT-LSTM模型的R^2值比其他2个模型更接近1。通过Arps模型进行产量预测的原理可以得知,时间跨度越大,所选代表井产量越大,Arps模型的预测精度就越低,甚至出现负数的情况。从另外2个指标RMSE和MAPE中也可以看出,Arps模型在数值上与LSTM模型有很大差距。Ⅱ类代表井的3种模型都可以在同一范围内进行比较分析,因此,将Ⅱ类代表井的各个评价指标的值都等量缩小至0~1之间,作出评价指标对比雷达图(图6)。在产量预测精度以及R^2三角形的最尖端都是AT-LSTM模型,RMSE和MAPE三角形的最尖端都是Arps模型。在有利于Arps模型的预测条件下,Arps模型仍然和LSTM模型具有不小的差距,因此,Arps模型并不适用于长7段页岩油的单井产量预测。

3.2.2 单井产量预测对比分析

根据已有生产数据,将长7段页岩油单井生产过程分为前期产量迅速提高期和后期产量递减期。分别对3类单井进行分阶段模拟预测,根据各分类单井的主要影响因素对产量变化做出合理解释,并有针对性地调整生

表6 不同分类井不同模型预测指标结果

Table 6 Prediction indicator results of different models for various well types

模型类别	Ⅰ类井				Ⅱ类井				Ⅲ类井			
	R^2	σ_{RMSE}	σ_{MAPE}	预测精度	R^2	σ_{RMSE}	σ_{MAPE}	预测精度	R^2	σ_{RMSE}	σ_{MAPE}	预测精度
Arps模型	-5.95	172.64	0.44	0.564 5	0.56	56.42	0.2	0.842 1	-0.87	126.29	0.447	0.553 7
LSTM模型	0.83	31.05	0.08	0.925 5	0.86	32.07	0.11	0.901 7	0.83	42.07	0.194	0.877 7
AT-LSTM模型	0.92	21.35	0.04	0.958 9	0.94	20.35	0.07	0.937 4	0.84	41.14	0.189	0.903 4

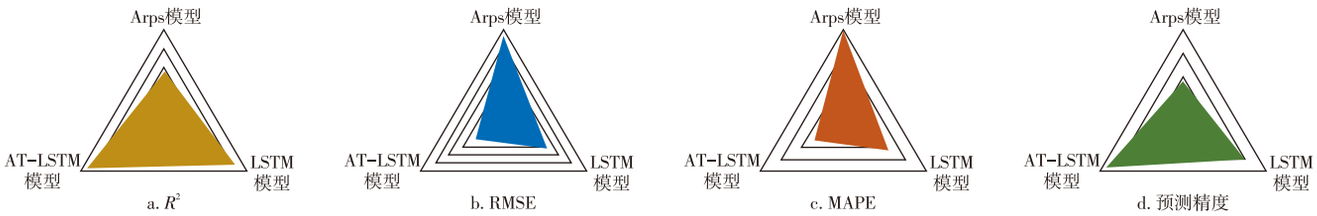


图6　Ⅱ类单井评价指标对比雷达图

Fig. 6　Radar charts of comparison evaluation indicators for class Ⅱ single wells

产措施。为了能够更有针对性地观察预测产量的变化，将各类单井代表井的连续单月产量制成折线图(图7)，这可以侧面反映出模型的稳定性。

Ⅰ类代表井周期长，前期产量迅速提高期长达6个月，并且初始产量高，后期产量递减期一直持续到第45个月(图7a)。在前期训练集预测中，AT-LSTM模型和LSTM模型均能较好地贴合实际产量曲线。但LSTM模型在后期训练集中产生了几个产量异常点，这是由于LSTM模型在处理长时间序列时容易把握不到重要的信息，导致产量预测出现偏差，而AT-LSTM模型则解决了这一问题，第29～33个月的红色曲线更加贴合蓝色曲线。在预测集中，AT-LSTM模型在趋势上与实际产量基本保持一致，而LSTM模型在第30～32个月的产量明显偏离实际产量，说明AT-LSTM模型能够在Ⅰ类单井产量预测中起到更好的效果。

Ⅱ类代表井周期较短，并且起始数据即为最高产量，仅有后期产量递减期(图7b)。从30个月中持续递减的产量可以看出，LSTM模型和AT-LSTM模型在总体趋势上没有出现较大偏差，但LSTM模型预测值大部分高于实际产量。针对没有前期产量提高期的问题，可以将输入的主要影响因素侧重于压裂段数和压裂簇数等后期调整生产措施时可以人为控制的变量，并加大其所分配的权重，因此AT-LSTM模型能够达到更加精确的模拟预测效果。

Ⅲ类代表井与Ⅰ类代表井类似，包含前期产量迅速提高期和后期产量递减期，但整体周期较短(图7c)。由于生产周期较短，LSTM模型在预测产量时并没有出现明显异常点，但整体产量偏高的问题依旧存在。针对整体周期较短的特点，将训练集的数据适当延长至前90%的单井产量数据，充分发挥注意力机制的作用，尽可能地减小误差。通过调整，后期测试集中AT-LSTM模型的产量曲线较前期训练集更加接近真实产量曲线。

为了更加清晰地看出注意力机制对LSTM模型的优化效果，针对3类单井分别绘制实际单月产量与预测产量的散点图(图8)。数据点越集中在斜对角线上，预测值越接近实际产量，模型的预测精度越高。其中，Ⅰ、Ⅱ类单井棕色点(引入注意力机制的LSTM模型)明显比绿色点(LSTM模型)更接近对角线，落在对角线附近的棕色点更多。至此，结合前面预测精度的分析可以得知，引入注意力机制的LSTM模型确实能够更加准确地预测出单井产量。另外可以看出Ⅰ类和Ⅲ类单井的单月产量跨度较大，Ⅱ类井单月产量较为集中。结合前文内容可以发现，Ⅱ类井代表井投产时已经达到最大产量，因此，Ⅱ类井预测产量曲线更加稳定，与实际产量的偏差也相对较小。选取最大产量时刻作为模型测试集的起始时刻有助于得到更为稳定的产量预测曲线，在单月产量的预测上也更为精确。

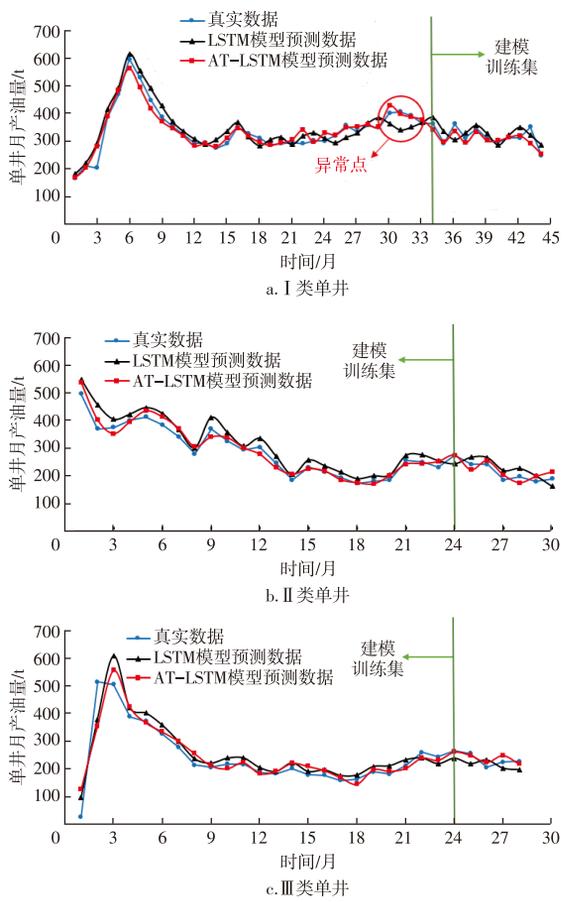


图7　Ⅰ-Ⅲ类单井不同模型的产量预测结果对比

Fig. 7　Comparison of production prediction results by different models for class Ⅰ-Ⅲ single wells

3.3　单井可控因素优化

通过对已筛选的197口长7段页岩油单井产量基础数据进行皮尔逊相关性分析，可以得到众多影响单井产

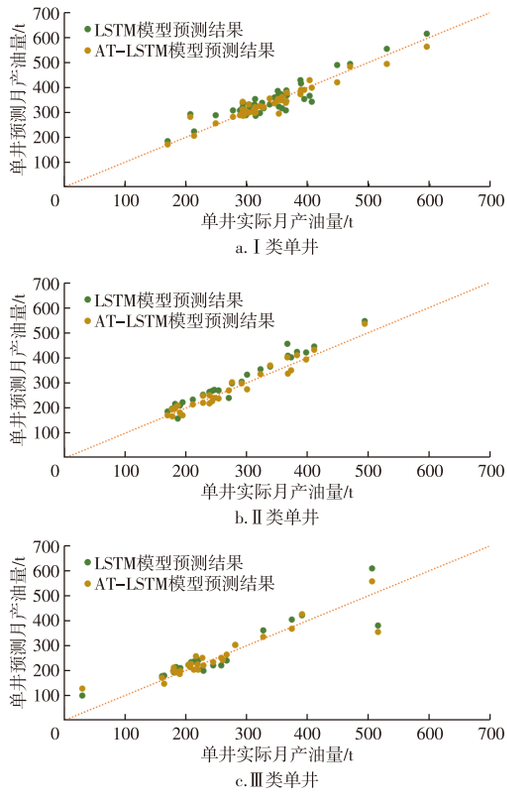


图8　Ⅰ-Ⅲ类单井产量预测结果

Fig. 8　Production prediction results for class Ⅰ-Ⅲ single wells

量的可控因素，以正式开采之前可调整的单井水平段长度为例，基于AT-LSTM智能模拟预测模型，通过差分进化算法以单井产量最大化为目标进行优化，并计算其理论最优值。单井产量采用单井日均产油量和单井总产油量进行优化研究。将水平段长度按照长度区间分段，由于单井水平段长度主要集中于1 500 m和2 000 m，故将这2个长度单独设置为一个水平段长度区间，剩余单井则按照每区间20口井左右进行设置，最终包含6个长度区间。通过优化模型统计出每一个区间的井数、平均水平段长度、平均日产量、平均总产量和单位平均日产量，如表7所示。表7中平均日产量最优为3.66 t，平均总产量最优为29 099.94 t，此时水平段长度为1 500 m，该区间共有67口单井。

上述分析以水平段长度为例，明确了关键可控因素的优化方向。基于同样的优化思想，研究进一步分析了入地总液量、加砂量与压裂段数的影响。模拟结果表明，这些因素同样存在一个明确的“协同最优区间”。当入地总液量处于33 000～39 000 m^3，加砂量处于4 800～6 500 m^3，压裂段数处于16～18段时，单井平均日产量与总产量能够同时达到最优。这一结果证实了多因素协同优化的可行性，为指导现场工程实践提供了量化的决策依据。

表7　水平段长度与产量的关系

Table 7　Relationship between horizontal section length and production

水平段长度/m	井数/口	平均水平段长度/m	平均日产量/t	平均总产量/t	单位平均日产量/(t/m)
[700, 1 500)	20	1 209.84	3.33	26 404.76	0.002 8
1 500	67	1 500.00	3.66	29 099.94	0.002 4
(1 500, 1 800)	29	1 620.25	3.45	27 265.97	0.002 1
[1 800, 2 000)	22	1 824.86	3.08	24 438.22	0.001 7
2 000	41	2 000.00	3.65	28 598.46	0.001 8
(2 000, 3 000]	21	2 368.48	3.56	28 366.94	0.001 5

4　结论

基于上述对长7段页岩油水平井产量智能模拟预测的研究，可得出以下3条核心结论：

1)基于数据驱动和特征加权FCM算法，长7段内筛选出的197口单井可分为3类。皮尔逊相关性分析表明，含油饱和度、压裂段数、加砂量、入地总液量和油层长度的平均相关性大小分别为0.51、0.54、0.40、0.52和0.38，该分析定性且定量地确定了3类单井产量的主要影响因素。

2)融合注意力机制与差分进化算法的LSTM智能产量预测模型，能够有效克服Arps模型和标准LSTM模型在长7段单井产量预测中的局限性，提升产量预测的精度和泛化能力。AT-LSTM模型在3类井产量的平均决定系数R^2、平均绝对百分误差MAPE、均方根误差RMSE和平均预测精度分别为0.900、0.067、18.720和0.933，均优于LSTM模型。

3)通过应用AT-LSTM模型对长7段单井主控因素进行优化，定量确定了水平段长度、入地总液量、加砂量和压裂段数的“协同最优区间”分别为1 500 m、33 000～39 000 m^3、4 800～6 500 m^3和16～18段，证实了通过多因素协同优化提升单井产量的可行性，为现场工程方案的优化提供了具体、量化的指导。

参 考 文 献

[1]　Liu C L, Wang Z L, Guo Z Q, et al. Enrichment and distribution of shale oil in the Cretaceous Qingshankou Formation, Songliao Basin, Northeast China[J]. Marine and Petroleum Geology, 2017, 86: 751-770.

[2]　Wang M, Li M, Li J B, et al. The key parameter of shale oil resource evaluation: Oil content[J]. Petroleum Science, 2022, 19(4): 1443-

1459.

[3] 金之钧, 张谦, 朱如凯, 等. 中国陆相页岩油分类及其意义[J]. 石油与天然气地质, 2023, 44(4): 801-819.
Jin Zhijun, Zhang Qian, Zhu Rukai, et al. Classification of lacustrine shale oil reservoirs in China and its significance[J]. Oil & Gas Geology, 2023, 44(4): 801-819.

[4] 赵文智, 卞从胜 李永新, 等. 陆相页岩油可动烃富集因素与古龙页岩油勘探潜力评价[J]. 石油勘探与开发, 2023, 50(3): 455-467.
Zhao Wenzhi, BIAN Congsheng, LI Yongxin, et al. Enrichment factors of movable hydrocarbons in lacustrine shale oil and exploration potential of shale oil in Gulong Sag, Songliao Basin, NE China[J]. Petroleum Exploration and Development, 2023, 50(3): 455-467.

[5] 何永宏, 薛婷, 李桢, 等. 鄂尔多斯盆地长7页岩油开发技术实践: 以庆城油田为例[J]. 石油勘探与开发, 2023, 50(6): 1245-1258.
He Yonghong, Xue Ting, Li Zhen, et al. Development technologies for Triassic Chang 7 shale oil in Ordos Basin: a case study of Qingcheng Oilfield, NW China[J]. Petroleum Exploration and Development, 2023, 50(6): 1245-1258.

[6] 陈劲松, 郭莉, 年静波, 等. 一种提高早期页岩油气单井产量预测合理性的方法[J]. 非常规油气, 2017, 4(6): 53-58.
Chen Jinsong, Guo Li, Nian Jingbo, et al. A method of improving reasonability on the shale oil and gas wells production prediction in the early stage[J]. Unconventional Oil & Gas, 2017, 4(6): 53-58.

[7] 陈劲松, 年静波, 韩洪宝, 等. 改进Arps递减模型早期产量预测再认识[J]. 非常规油气, 2019, 6(1): 75-80.
Chen Jinsong, Nian Jingbo, Han Hongbao, et al. Improve production prediction rationality of the modified Arps decline model in the early shale oil and gas wells[J]. Unconventional Oil & Gas, 2019, 6(1): 75-80.

[8] 陈劲松, 曹健志, 韩洪宝, 等. 页岩油气井常用产量预测模型适应性分析[J]. 非常规油气, 2019, 6(3): 48-57.
Chen Jinsong, Cao Jianzhi, Han Hongbao, et al. Adaptability analysis of commonly used production prediction models for shale oil and gas well[J]. Unconventional Oil & Gas, 2019, 6(3): 48-57.

[9] 董鹏, 陈志明, 于伟. 压裂后页岩油藏多裂缝直井产能模型: 以鄂尔多斯盆地页岩油井为例[J]. 大庆石油地质与开发, 2022, 41(1): 155-165.
Dong Peng, Chen Zhiming, Yu Wei. Study on productivity model for multiple-fracture vertical well in shale oil reservoirs after fractured: a case of shale oil wells in Ordos Basin[J]. Petroleum Geology & Oilfield Development in Daqing, 2022, 41(1): 155-165.

[10] 杨洋, 程悦菲, 谯英, 等. 基于时序动态分析的油井产量预测研究[J]. 西南石油大学学报(自然科学版), 2020, 42(6): 82-88.
Yang Yang, Cheng Yuefei, Qiao Ying, et al. Research on oil well production prediction based on time-series dynamic analysis[J]. Journal of Southwest Petroleum University(Science & Technology Edition), 2020, 42(6): 82-88.

[11] 刘阳, 陈思彤, 向幸运, 等. 基于多维时序LSTM的超深井机械钻速预测方法[J]. 西南石油大学学报(自然科学版), 2025, 47(5): 121-133.
Liu Yang, Chen Sitong, Xiang Xingyun, et al. Prediction method of mechanical drilling rate for ultra-deep wells based on multi-dimensional time-series LSTM[J]. Journal of Southwest Petroleum University(Science & Technology Edition), 2025, 47(5): 121-133.

[12] 闵超, 郭星, 华青, 等. 基于FCMFS特征选择算法的煤层气压裂效果预测[J]. 西南石油大学学报(自然科学版), 2025, 47(2): 95-104.
Min Chao, Guo Xing, Hua Qing, et al. Coalbed methane fracturing effect prediction based on FCMFS feature selection algorithm[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2025, 47(2): 95-104.

[13] 官文洁, 彭小龙, 朱苏阳, 等. 基于改进LSTM神经网络的加密井产能预测研究: 以川南中深层页岩气为例[J]. 油气藏评价与开发, 2025, 15(3): 479-487.
Guan Wenjie, Peng Xiaolong, Zhu Suyang, et al. Research on productivity prediction method of infilling well based on improved LSTM neural network: a case study of the middle-deep shale gas in South Sichuan[J]. Petroleum Reservoir Evaluation and Development, 2025, 15(3): 479-487.

[14] Bezdek J C, Ehrlich R, Full W. FCM: the fuzzy c-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2/3): 191-203.

[15] Gao X B, Xie W X. Advances in theory and applications of fuzzy clustering[J]. Chinese Science Bulletin, 2000, 45(11): 961-970.

[16] Sun L R, Mao H F, Zheng C, et al. Interval-valued functional clustering based on the improved Euclidean distance with application to air quality index[J]. Applied Mathematical Modelling, 2023, 123: 627-643.

[17] Munandar T A, Sumiati S, Rosalina V. Pattern of symptom correlation on type of heart disease using approach of Pearson correlation coefficient[J]. IOP Conference Series: Materials Science and Engineering, 2020, 830(2): 022086.

[18] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention[C]// Advances in Neural Information Processing Systems 27, 2014.

[19] Storn R, Price K. Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359.

[20] 张春晓. 基于深度长短期记忆神经网络的油气井产量预测优化方法[J]. 石油化工应用, 2023, 42(11): 28-31.
Zhang Chunxiao. Optimization method for oil and gas well production prediction based on deep long short-term memory neural network[J]. Petrochemical Industry Application, 2023, 42(11): 28-31.

[21] 王洪亮, 林霞, 蒋丽维, 等. 基于聚类及长短时记忆神经网络预测油田产量[J]. 石油科学通报, 2024, 9(1): 62-72.
Wang Hongliang, Lin Xia, Jiang Liwei, et al. An oilfield production prediction method based on clustering and long short-term memory neural network[J]. Petroleum Science Bulletin, 2024, 9(1): 62-72.

[22] Das S, Suganthan P N. Differential evolution: a survey of the state-of-the-art[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 4-31.

(编辑 顾丽媛)