

拟合函数—神经网络协同的页岩气井产能预测模型

胡晓东*, 涂志勇, 罗英浩, 周福建, 李宇娇, 刘健, 易普康

中国石油大学(北京)人工智能学院, 北京 102249

* 通信作者, huxiaodong@cup.edu.cn

收稿日期: 2021-11-01

中石油战略合作项目“物探、测井、钻完井人工智能理论与应用场景关键技术研究(No. ZLZX2020-03)”资助

摘要 气井产能预测是气田开发的一项重要任务。而页岩气生产受地质和生产中诸多因素影响, 非线性特征强, 传统基于机理的产能预测方法难以综合且准确表征多维度和多结构类型的产能影响因素, 难以快速求解页岩气压裂后生产动态。针对这一问题, 基于LSTM和DNN神经网络, 本文提出了一种新型的拟合函数—神经网络协同的页岩气井动态产能预测模型。首先, 通过数据维度重组, 将目标井前期的产量、压力等时序参数与用液强度、加砂强度、总含气量、脆性矿物含量等静态产能控制参数混合构建数据集, 以实现目标井后期生产曲线的预测。其次, 基于现场真实日产气量数据, 利用Arps产能曲线拟合模型对同区块的邻井产能数据进行特征筛选, 以间接加入含产能递减规律的弱物理约束; 基于实际工况下单日生产时间与产量的强相关性, 于神经网络模型内部加入强物理约束, 进而提高本文模型的产能时间序列预测精度和局部稳定性。基于本文模型, 对我国某页岩气区块进行未来产能曲线预测, 并对预测结果进行k折交叉验证。其中, 分别讨论了神经网络模型参数、产能控制参数和时间步长等因素对于模型精度的影响。结果显示, 本文模型具有较高的准确率。在邻井生产数据样本较少情况下, 该模型仍可以通过用液强度等静态产能控制参数和目标井前期生产、压力曲线捕捉较多的生产特征。本文为老井压裂效果评价和新井生产参数优化具有一定指导意义。

关键词 LSTM; 物理约束; 动态产能预测; 页岩气; 机器学习

Shale gas well productivity prediction model with fitted function-neural network cooperation

HU Xiaodong, TU Zhiyong, LUO Yinghao, ZHOU Fujian, LI Yujiao, LIU jian, YI Pukang

College of Artificial Intelligence, China University of Petroleum-Beijing, Beijing 102249, China

Abstract Gas well productivity prediction is an important task in gas field development. In contrast, shale gas production is influenced by many factors in geology and production with strong nonlinear characteristics. Traditional mechanism-based productivity prediction methods are difficult to comprehensively and accurately characterize multi-dimensional and multi-structural types of productivity influencing factors, and it is difficult to quickly solve the production dynamics after shale gas fracturing. To address this problem, based on LSTM and DNN, a novel fitting function-neural network synergistic model for dynamic production productivity prediction of shale gas wells was proposed in this paper. Firstly, the data set was constructed by reorganizing the data dimensions, mixing the time-series parameters such as production and pressure in the early stage of the target well with the

引用格式: 胡晓东, 涂志勇, 罗英浩, 周福建, 李宇娇, 刘健, 易普康. 拟合函数—神经网络协同的页岩气井产能预测模型. 石油科学通报, 2022, 03: 394-405

HU Xiaodong, TU Zhiyong, LUO Yinghao, ZHOU Fujian, LI Yujiao, LIU jian, YI Pukang. Shale gas well productivity prediction model with fitted function-neural network cooperation. Petroleum Science Bulletin, 2022, 03: 394-405. doi: 10.3969/j.issn.2096-1693.2022.03.034

static productivity control parameters such as fluid intensity, sand addition intensity, total gas content, brittle mineral content, etc., in order to achieve the prediction of the production curve in the late stage of the target well. Second, based on the real daily gas production data in the field, the Arps productivity curve fitting model was used to filter the productivity data of neighboring wells in the same block to indirectly add a weak physical constraint containing the law of decreasing productivity; based on the strong correlation between single-day production time and production under actual working conditions, a strong physical constraint was added inside the neural network model to improve the productivity time series prediction accuracy and local stability of this model. This improves the prediction accuracy and local stability of the model. Based on this model, a shale gas block in China was predicted to have a future production curve, and the prediction results were cross-validated by k-fold Method. Among them, the effects of neural network model parameters, productivity control parameters and time step on the model accuracy were discussed separately. The results show that the model in this paper has a high accuracy rate. With a small sample of production data from neighboring wells, the model can still capture more production characteristics by using static capacity control parameters such as fluid intensity and pre-production and pressure profiles of the target wells. This study results in this paper provide some guidance for the evaluation of fracturing effect of old wells and the optimization of production parameters of new wells.

Keywords LSTM; physical constraint; dynamic productivity prediction; shale gas; machine learning

doi: 10.3969/j.issn.2096-1693.2022.03.034

0 前言

页岩气压裂后产能预测对于水平井压裂效果评价和生产参数优化具有重要意义。传统的压后产能预测方法包含解析法^[1-3]、半解析法^[4-6]、数值模拟法^[7-8]和递减曲线拟合法^[9-10]等。传统解析和数值模拟方法存在计算复杂、多物理场耦合困难等问题。而基于递减曲线等传统经验模型只能粗略描述地递减特征,难以准确而全面地表征复杂条件下的生产动态。

20世纪以来,许多学者尝试将基于数据驱动的机器学习算法应用到油气领域^[11-13]。而针对气井产能预测的数据驱动算法是其中一种重要应用场景^[14]。数据驱动方法能够避免物理机理分析方法构建的复杂性和多解性^[15]。具体而言,单井产能预测问题的输入和预测对象均具有时间序列维度上连续性和自相关性。然而,地质和生产参数与产能随时间的变化具有强非线性关系^[16]。针对这一问题,一些学者应用支持向量机^[16-17]、粒子群^[18]、BP神经网络^[19-20](Back Propagation Neural Network)、循环神经网络^[21]和LSTM神经网络^[22-23](Long-Short Term Memory Neural Network)等算法的变形或耦合模型构建产能预测模型,以更好地捕获地质和生产两方面中多个控制参数(如井斜、垂深、孔隙度、排量、射孔簇数、支撑剂等)对生产曲线的非线性关系,从而提高产能预测的准确率。这些机器学习算法能够简化产能预测模型的构建和动态求解速度,但在产能预测模型优选、小样本训练、非结构化数据处理和主控因素分析等方面仍存在挑战。首先,产能预测的关键预测对象是气井正式投产后的未来产量动态变化,其影响因素不仅包括地质和生产过

程中与时间无关的静态参数,还包括气井前期生产的产能时序数据。如何在时序序列预测模型中考虑静态参数是产能预测神经网络模型构建的关键。此外,气井产能预测模型的现场数据样本较少,且一些样本中存在异常产能趋势,这会造成模型训练的过拟合和泛化能力差等问题。针对以上问题,如何对神经网络模型结构和预测流程进行适应性修正是目前亟需解决的问题。

本文基于LSTM和DNN(Deep Neural Network)神经网络,建立了拟合函数—神经网络协同的页岩气井动态产能预测模型。首先,针对非结构化数据,本文利用维度扩展的方法表征了静态数据的时间不变性,并将前期日产气量、平均套管压力共同组成的二维时序数据与用液强度、加砂强度、总含气量、脆性矿物含量等静态产能控制参数混合构建数据集,以实现时序—非时序混合数据的时间序列预测。其次,本文将传统产能递减模型作为LSTM神经网络输入层的“弱物理约束”,即在特征工程阶段通过产能递减模型的拟合函数对所有准备训练的井样本数据进行预筛选。其中,Arps产能递减微分方程是Arps^[9,24]提出的基于概率统计的产能预测模型,陈元千和郝明强^[25]在此基础上推导出Arps的指数递减、双曲递减和调和递减等3种不同递减系数下的产能模型。本文基于Arps的3类递减拟合函数,通过谢军等人^[26]提出的筛选法,以最小均方误差(mean squared error, MSE)为标准自动优选递减模型,对我国某页岩区块中的各单井产能数据进行递减曲线拟合,去除不符合产能递减规律的无效样本。同时,将实际工况下单日生产时间与产量的比例关系作为DNN神经网络输出层的“强物理约束”,

即在神经网络的最后一个全连接层后对输出结果按实际单日产时间进行缩放,以避免实际工况下由于临时停泵等人为因素造成的产能曲线异常点对模型反向优化过程的影响。基于以上两类“物理约束”,本文模型既能捕获产能递减的共性规律,也能避免人为因素造成的特征干扰,解决了模型的无物理规律、过拟合的问题,从而保证了预测结果的稳定性。

最后,对筛选后的页岩气井的动态产能预测进行k折(k-fold)交叉验证,并分别讨论了神经网络模型参数、产能控制参数和时间步长等对于本文模型误差的影响。结果显示,LSTM+DNN神经网络可以通过用液强度、加砂强度、总含气量、脆性矿物含量等静态产能控制参数,前期生产、压力曲线和单日产时间序列较快速准确地捕捉的后期生产特征。此外,通过拟合函数和神经网络的协同,在不增加样本规模的条件下模型能较准确和稳定地预测后期产能递减趋势。本文模型对于页岩气动态产能预测具有更好的预测能力,对老井压裂效果评价和新井生产参数优化具有一定指导意义。

1 方法与模型

1.1 LSTM神经网络原理

长短期记忆神经网络(LSTM)主体结构类似于RNN(Recurrent Neural Network)的链状重复网络结构,主要区别在于将RNN模型中隐藏层中的神经元替换为LSTM特有的结构单元。其典型的结构单元(见图1)包括“遗忘门”(forget gate)、“输入门”(input gate)、“输出门”(output gate)^[27]。

基于以上3个组成部分,LSTM网络结构单元典型的前向计算流程如下^[27-29]:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

式中, f_t 为“遗忘门”的输出值(0~1间的数字),决定了上一步处理器状态(cell state)应被继续传递的部分; i_t 为“输入门”的输出值,决定了tanh类激活函数计算结果中应被保留的部分; c_t 为tanh类激活函数的计算结

果,表征处理器状态的值; o_t 为“输出门”的输出值; h_t 为整个隐藏层的输出值; W 表示各单元之间的权重系数矩阵; b 表示各单元对应的偏置项; σ 表示Sigmoid激活函数。

1.2 基于LSTM+DNN的动态产能预测模型

动态产能预测模型基于地质及工程等静态参数,产能、平均套管压力等时序参数和单日产时间等强约束控制参数共同构建。现场数据包括50口页岩气井。其中,实测递减阶段的产能数据满足模型构建所需最小时间序列的井样本共24口。整理出供优选的地质参数8个(杨氏模量、总有机碳含量、最大水平主应力、最小水平主应力、总含气量、孔隙度、泊松比和脆性矿物含量)、工程参数6个(用液强度、加砂强度、平均泵压、平均排量、段长和簇间距)。结合灰色关联法和熵权法,确定各地质、工程参数与产能的相关度^[30]。为确保模型预测的稳定性,从地质、工程参数中各选取与产能相关度最高的前2个参数作为静态产能控制参数,即用液强度、加砂强度、总含气量、脆性矿物含量。

由于静态产能控制参数无法直接和生产数据共同构建时间序列数据集,因此需要进行维度扩展处理。具体而言,模型训练包括多口井的数据,其中每口井具有与时间相关的产能、平均套管压力数据和特定且

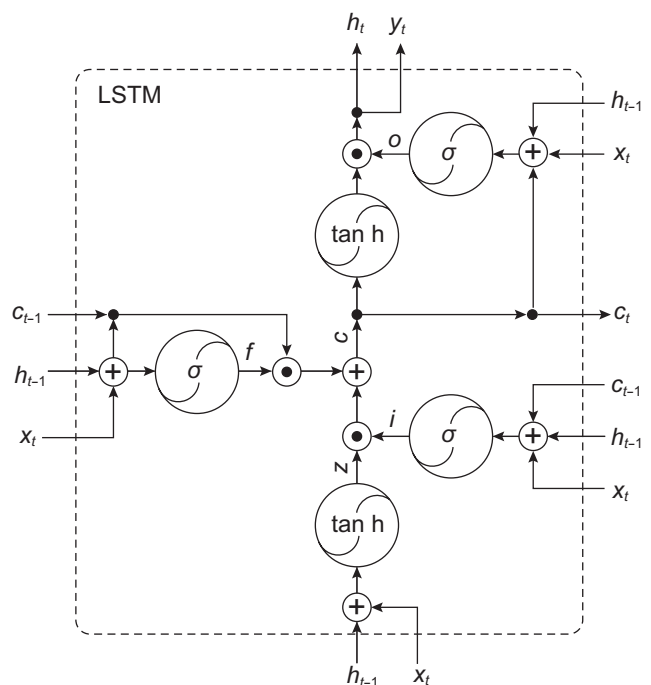


图1 LSTM网络结构单元^[28]

Fig. 1 LSTM network structure unit^[28]

唯一的一组时间无关的静态产能控制参数。为了同时训练并提取产能控制参数和前期产能、平均套管压力数据对于后期产能数据的非线性规律,将每个时间无关的产能控制参数视为在该井后期产能曲线范围内与时间相关,并重复该范围内的静态控制参数,扩展到后期产能曲线的同一时间维度。当后期产能曲线在时间维度上的每个时间点具有对应的一组产能控制参数值,则可将所有井前期产能、平均套管压力数据和后期产能控制参数(包括单日实际生产时间)作为输入特征向量,后期产能数据作为输出特征向量,从而生成非结构化的高维时间序列数据集。最后,基于LSTM和DNN神经网络,设计了多变量动态产能预测模型(如图2)。模型训练和迭代过程包括:(1)对产能样本进行特征筛选,并构建非结构化的多变量时间序列数据集;(2)标准化数据集,划分训练集和测试集;(3)基于LSTM和DNN神经网络前向计算方法计算模型输出值;(4)基于网络层级和时间反向传播,计算模型输出值和实际值的误差项;(5)计算各对应权重的梯度;(6)基于梯度的优化算法更新权重;(7)重复迭代,确定最优的产能预测结果,并将结果反标准化。

产能模型包含6个产能控制参数(平均套管压力、用液强度、加砂强度、总含气量、脆性矿物含量和单日实际生产时间),预实现通过某井前 n 个步长的产能数据预测后 m 个步长的产能数据。其中,模型主要设计为2个LSTM层和3层全连接层,分别通过3个输入接口传入所有产能控制参数和前期产能时序数据。模型对数据本身量级非常敏感,因此输入前需利用min-max标准化将所有数据映射到0~1之间,以消除参数之间的量纲影响。具体而言,首先,针对模型中前2个LSTM层,主要用于捕获产能递减趋势特征,需于第1个LSTM层输入由前期产能时序数据和对应平均套管压力时序数据组成的二维特征向量,即

$$X_1 = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_n \\ x_{1,1} & x_{1,2} & \cdots & x_{1,n} \end{bmatrix} \quad (6)$$

其中 Q_n 为第 n 时刻(步)的产能数据, $x_{1,n}$ 为第 n 时刻(步)的平均套管压力, X_1 表示第一组输入特征向量。

前2个LSTM层将输出一组一维的中间变量,其长度与预测生产曲线的时间序列步长(m)一致。同时,将用液强度、加砂强度、总含气量、脆性矿物含量维度扩展为预测生产曲线的时间序列步长(m),并与LSTM层输出的中间变量共同组成一组新的五维特征向量,即

$$X_2 = \begin{bmatrix} x_{2,n+1} & x_{2,n+2} & \cdots & x_{2,n+m} \\ x_{3,n+1} & x_{3,n+2} & \cdots & x_{3,n+m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{6,n+1} & x_{6,n+2} & \cdots & x_{6,n+m} \end{bmatrix} \quad (7)$$

其中, $x_{2,n+m}$ 为LSTM层输出的中间变量, $x_{3,n+m} \sim x_{6,n+m}$ 分别为维度扩展后的用液强度、加砂强度、总含气量、脆性矿物含量数据, X_2 表示第二组输入特征向量。

通过3个全连接层后将输出得到第 n 时刻至第 $n+m$ 时刻间的产能数据。然后,输入同时间段内各个时刻的单日实际生产时间,作为第三组输入特征向量,即

$$X_3 = [t_{n+1} \quad t_{n+2} \quad \cdots \quad t_{n+m}] \quad (8)$$

基于单日实际生产时间与对应单日产能数据的等比例关系,修正3个全连接层输出的产能,修正公式如下

$$Q'_{n+m} = \frac{t_{n+m}}{24} Q_{n+m} \quad (9)$$

其中, Q_{n+m} 和 Q'_{n+m} 分别为修正前后的无因次产能预测值,需要通过反标准化确定最终的预测产量。

根据单井产能预测的实际条件,在训练过程中需要以井为单位分批构建时间序列进行模型训练,每口井重组10个样本(当单井实测时间序列不足划分10个样本时,取其能够划分的多样本数)。其中,从各井产能递减阶段的起始位置开始,向后取点作为样本时间序列的初始时刻 T ,并取 $T \sim T+n$ 时刻的生产曲线作为已知时间序列, $T+n+1 \sim T+n+m$ 时刻的生产曲线作为预测目标。

2 结果与讨论

2.1 基于Arps产能递减模型的特征筛选

对于动态产能预测模型,当训练集的大量井样本中出现过多异常的无效样本,会造成神经网络特征提取过程中计算出错误的权重,反而会降低模型预测的精度。因此,在LSTM动态产能预测模型训练前需要进行一定预处理。除了对数据标准化,还需要筛选去除产量曲线明显异常的样本,以提高动态产能预测模型的特征提取的准确率。针对我国某页岩区块的单井产能曲线预测进行特征筛选。筛选对象共包含24口同区块页岩气井的真实产能数据。本文以Arps产能递减模型的拟合曲线作为评判标准筛选出12口符合产能递减规律的井样本(图3)。其中,Arps产能递减模型的一般微分方程表示为^[25]

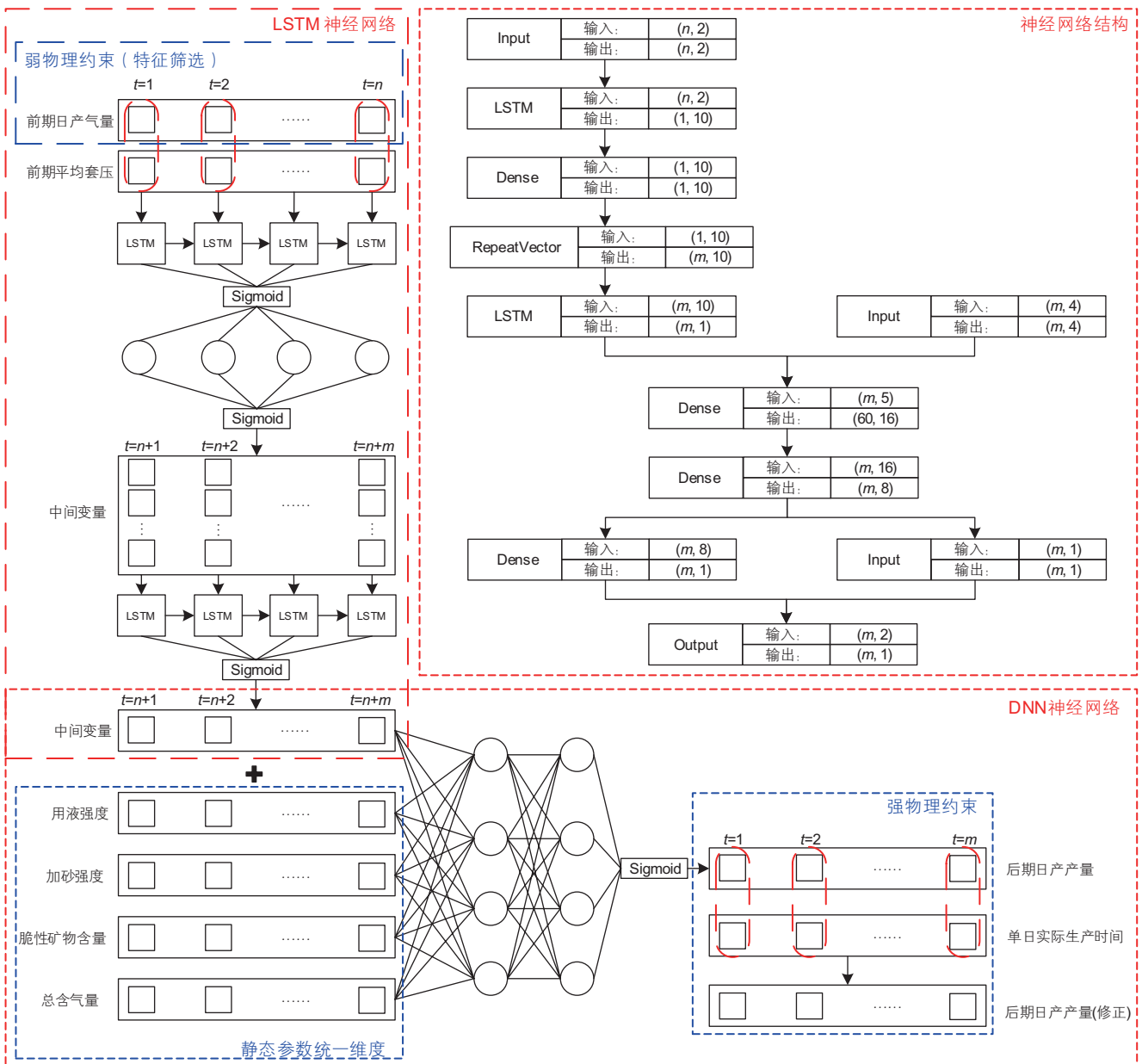


图 2 LSTM产能预测模型结构
Fig. 2 LSTM productivity prediction model structure

$$-\frac{dq}{dt} = nt + D_i^{-1} \quad (10)$$

其中, q 为产量($10^4 \text{ m}^3/\text{天}$), t 为时间(天), n 为递减指数, D_i 为初始产量($10^4 \text{ m}^3/\text{天}$)。

根据递减指数 n 的不同, Arps微分方程可进一步表示为

$$q = \begin{cases} \frac{q_i}{(1+nD_i t)^{1/n}}, 0 < n < 1, \text{双曲递减模型} \\ q_i e^{-D_i t}, n = 0, \text{指数递减模型} \\ \frac{q_i}{1+nD_i t}, n = 1, \text{调和递减模型} \end{cases} \quad (11)$$

分别利用3种模型拟合递减曲线, 并比较不同拟合曲线的MSE, 以最小MSE为标准自动优选递减模型拟合曲线。

基于特征筛选后的井样本, 以天为单位对各井日产气量数据进行时间序列化, 并通过k-fold交叉验证评价模型精度。具体而言, 轮流将每口井作为测试集、其他11口井作为训练集进行模型训练和产能预测, 并计算对应的MSE, 最后将12次测试得到的均方误差取平均, 获得最终用于评价模型精度的误差值。当误差值越大, 表明模型预测精度越低。

以k-fold交叉验证所得的平均误差值为基础, 通

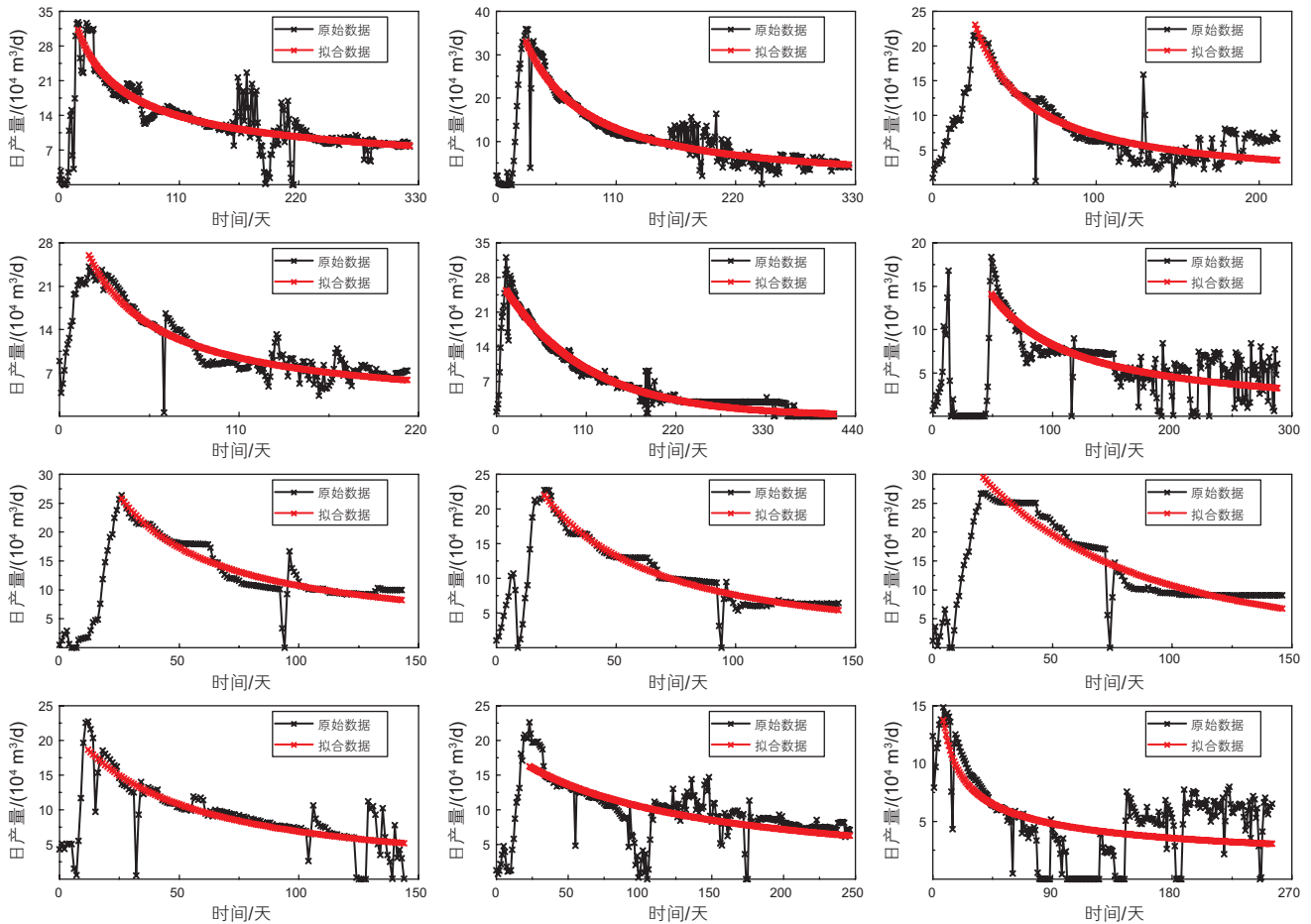


图3 基于Arps产能递减模型的训练及测试井样本筛选

Fig. 3 Training and test well sample selection based on Arps productivity model

通过对神经网络模型参数的优选确定最优的动态产能预测模型设计方案,最后讨论了静态产能控制参数和时间步长对模型精度的影响。

2.2 神经网络模型参数的影响规律

神经网络模型参数包括激活函数和梯度优化函数等。激活函数决定了神经网络前向计算时各层神经元向下一层传递的输出。除了linear激活函数,其他激活函数均为非线性函数,能够强化网络的非线性学习能力。梯度优化函数是神经网络在反向传播计算误差项时引导损失函数梯度下降的优化函数,即令损失函数的参数沿某一方向更新,使损失函数值逼近全局最小。梯度优化函数决定了损失函数的优化方向和步长。因此,针对本文模型优选激活函数和梯度优化函数能够提高模型的收敛速度和预测精度。

对于模型中LSTM层的2层输出的激活函数,分别取ELU、SELU、tanh、Sigmoid和linear激活函数进

行交叉验证实验,确定各激活函数及其组合下的模型精度如表1所示。结果显示,ELU、tanh和Sigmoid函数的模型预测效果相对较好。综合考虑各函数条件下的模型对于各井的预测误差和平均误差,优选Sigmoid函数作为模型的激活函数。其中Sigmoid函数表示为

$$\text{sigmoid}(x) = \frac{1}{1 + e^x} \quad (12)$$

对于模型的梯度优化函数。令LSTM层的激活函数为ELU和tanh函数,分别取梯度优化函数为SGD^[31]、RMSprop^[32]、Adagrad^[33]、Adadelta、Adam、Adamax和Nadam^[34]进行交叉验证实验,确定各梯度优化函数下的模型误差如表2所示。结果显示,Nadam函数的模型预测效果最好。其梯度计算公式如下^[34]

$$\begin{cases} g_t = \nabla_{\theta_{t-1}} f_t(\theta_{t-1}) \\ \theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{n_t + \varepsilon}} m_t \end{cases} \quad (13)$$

式中, g_t 为平均梯度, θ_t 为目标更新参数, n_t 为梯度的

二阶矩估计值, m_l 为梯度一阶矩, ϵ 为初始常数, α 为动量系数。

最后, 以Sigmoid作为激活函数, Nadam作为梯度优化函数训练动态产能预测模型, 预测结果见图4。交叉验证结果显示, 基于Sigmoid激活函数和Nadam梯度优化函数的动态产能预测模型的平均误差

为2.0001, 其中大部分井具有较好的预测效果。对于实际对于同一区块的具有相似产能递减规律的气井, 通过对多个老井的训练, 得到的模型能够基于新井前期部分产能曲线和用液强度、加砂强度、总含气量、脆性矿物含量等产能控制参数共同预测未来产能递减曲线。神经网络模型结构包含单日实际生产时间的强约束, 有效降低了实际工况下大量异常产能数据点对

表1 不同激活函数下的模型预测精度

Table 1 Prediction accuracy of models with different activation functions

激活函数	模型误差(MSE)
ELU	4.3236
SELU	5.9634
tanh	4.9023
Sigmoid	3.7201
linear	5.1589
ELU+tanh	4.3377
ELU+Sigmoid	5.3463
tanh+Sigmoid	5.3472

表2 不同梯度优化函数下的模型的误差对比

Table 2 Error comparison of models with different gradient optimization functions

梯度优化函数	模型误差(MSE)
SGD	4.0779
RMSprop	3.7201
Adagrad	3.7047
Adadelta	14.1435
Adam	4.7536
Adamax	4.1634
NAdam	2.0001

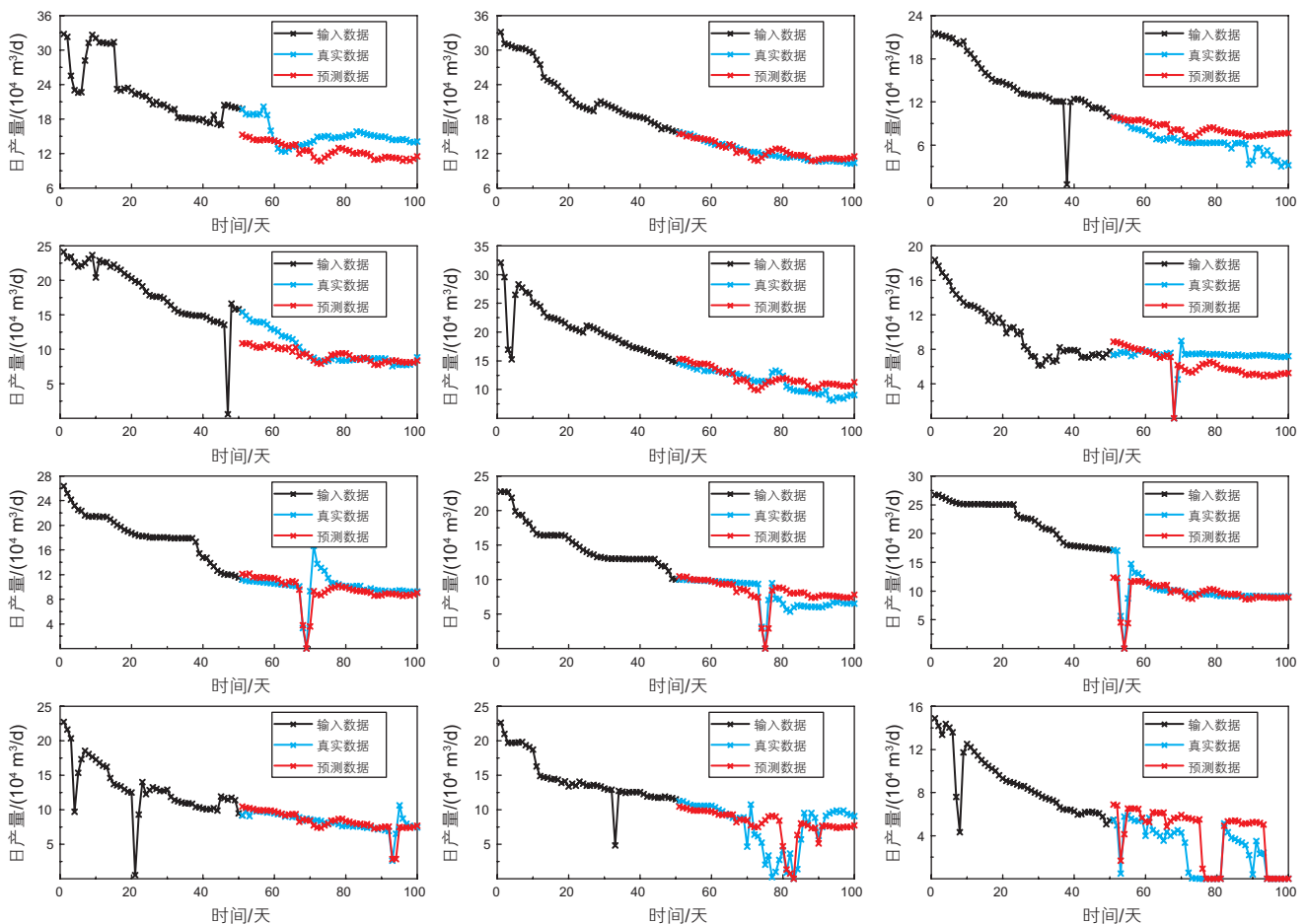


图4 动态产能预测模型预测结果

Fig. 4 Dynamic productivity forecasting model prediction results

产能趋势的影响。

2.3 静态产能控制参数的影响规律

本小节主要分析了静态产能控制参数对产能预测模型的影响。首先,保持神经网络模型结构不变,去除所有静态参数的输入接口,即第二组输入特征向量 X_2 变为前2个LSTM层输出的一维中间变量。以2.2中确定的神经网络超参数重新训练模型,并与考虑静态控制参数的原模型预测结果进行对比(图5)。

结果显示,不含静态产能控制参数(用液强度、加砂强度、总含气量和脆性矿物含量)的模型预测结果整体偏高。对于同一口井,其静态产能控制参数不变。当通过2层LSTM神经网络确定后期产能趋势预测结果后,静态产能控制参数主要通过DNN神经网络对后期各天的产能预测值进行校正,并修正对应模型权重。极端条件下,当产能控制参数数值变化对产能影响较大,其校正后的产能预测值将偏离LSTM确定的整体产能趋势。例如单日实际生产时间与单日产能具有强烈的正比例关系,图5中第90~100天内单日实际生产时间由24小时骤减为9

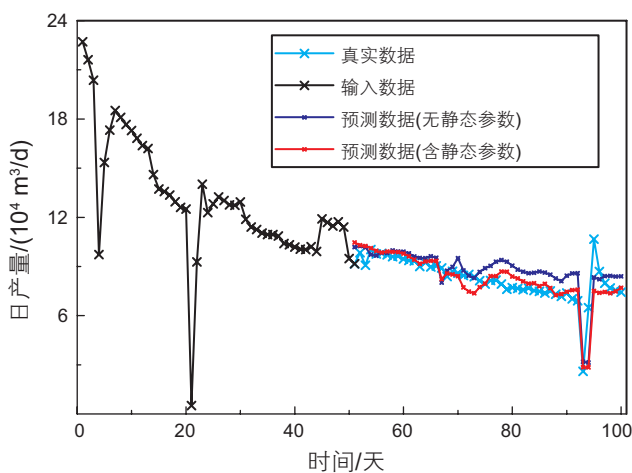


图5 考虑与不考虑静态产能控制参数的10井产能预测结果
Fig. 5 Productivity prediction results for well-10 with and without static productivity control parameters

小时,其产能预测值同样远小于趋势预测结果。固定其他产能参数不变,将10井的加砂强度分别设置为1.415、1.978和2.542(所有样本的加砂强度范围为1.415~2.542 t/m),设置三组实验(图6)。结果显示,随着加砂强度的增加,后期的平均日产气量同样增加。

2.4 输入时间步长和预测时间步长的影响规律

本小节分别分析了输入时间步长和预测时间步长对动态产能预测模型预测精度的影响。其中预测模型参数设置由2.2小节的激活函数和优化函数优选结果确定。首先固定预测时间步长为90天,分别设置输入时间步长为40,50,60和70天,对12口井进行k-fold交叉验证,各输入时间步长条件下的各井预测的平均误差结果如表3所示,其中2井的预测结果如图7所示。

固定输入时间步长为50天,分别设置预测时间步长为30、50、70和90天,对12口井进行k-fold交叉验证,各预测时间步长条件下的各井预测的平均误差结果如表4所示,其中4井的预测结果如图8所示。输入时间步长为50天,设置预测时间输出步长为200

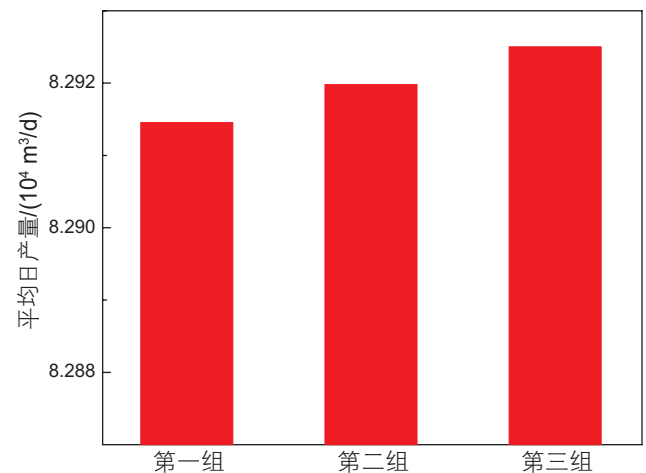


图6 不同加砂强度条件下的10井平均产能预测结果
Fig. 6 Average productivity prediction results of well-10 with different sand addition strength conditions

表3 输入时间步长为40,50,60和70天的5井产能预测模型的误差对比

Table 3 Error comparison of productivity prediction models with input time steps of 40, 50, 60 and 70 days

输入时间步长/天	模型误差(MSE)
40	1.9003
50	1.9042
60	2.2006
70	2.7232

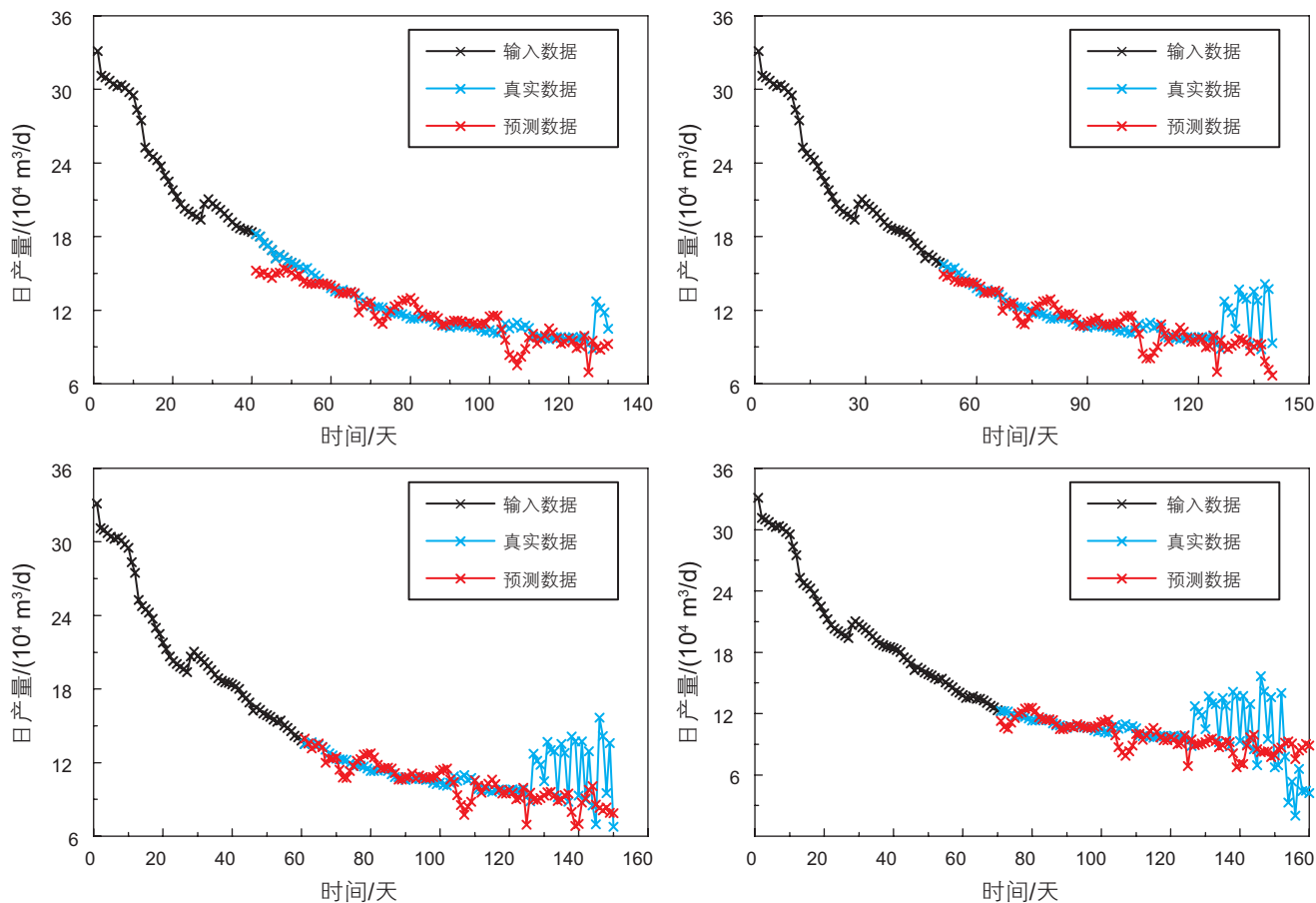


图7 输入时间步长分别为40、50、60和70天的2井产能预测结果

Fig. 7 well-2 productivity prediction results for input time steps of 40, 50, 60 and 70 days, respectively

天, 其中2井的预测结果如图9所示。

结果显示, 在40~70天范围内, 输入时间步长对产能预测精度具有较大影响。当已知的时间序列数据越多, 模型精度反而越低。其中, 前期异常衰减的产能数据点对预测结果影响较大。模型中单日实际生产时间主要约束未来产能递减趋势中的异常衰减点。随着输入时间步长的增加, 各井前期存在的衰减点变多, 且避开了单日实际生产时间的强约束, 削弱了产能趋势特征。其次, 预测时间步长的增加并没有明显降低模型预测的精度。由于数据预处理过程对产能曲线样

本进行了特征筛选, 各样本的产能曲线均满足一般递减规律。基于序列到序列的LSTM神经网络, 能够有效建立前期产能、平均套管压力曲线时间序列和后期递减曲线时间序列的特征关系。同时, 通过多个静态产能控制参数的权重校正和单日时间生产时间的强约束, 进一步提高了后期产能曲线的局部预测精度。最后, 通过2井的200天产能预测结果, 进一步说明动态产能预测模型对于更长期产能曲线预测具有较好的适用性。

表4 预测时间步长为30、50、70和90天的产能模型的误差对比

Table 4 Error comparison of productivity prediction models with forecast time steps of 30, 50, 70 and 90 days

预测时间步长/天	模型误差(MSE)
30	2.1240
50	2.0001
70	2.2477
90	1.9042

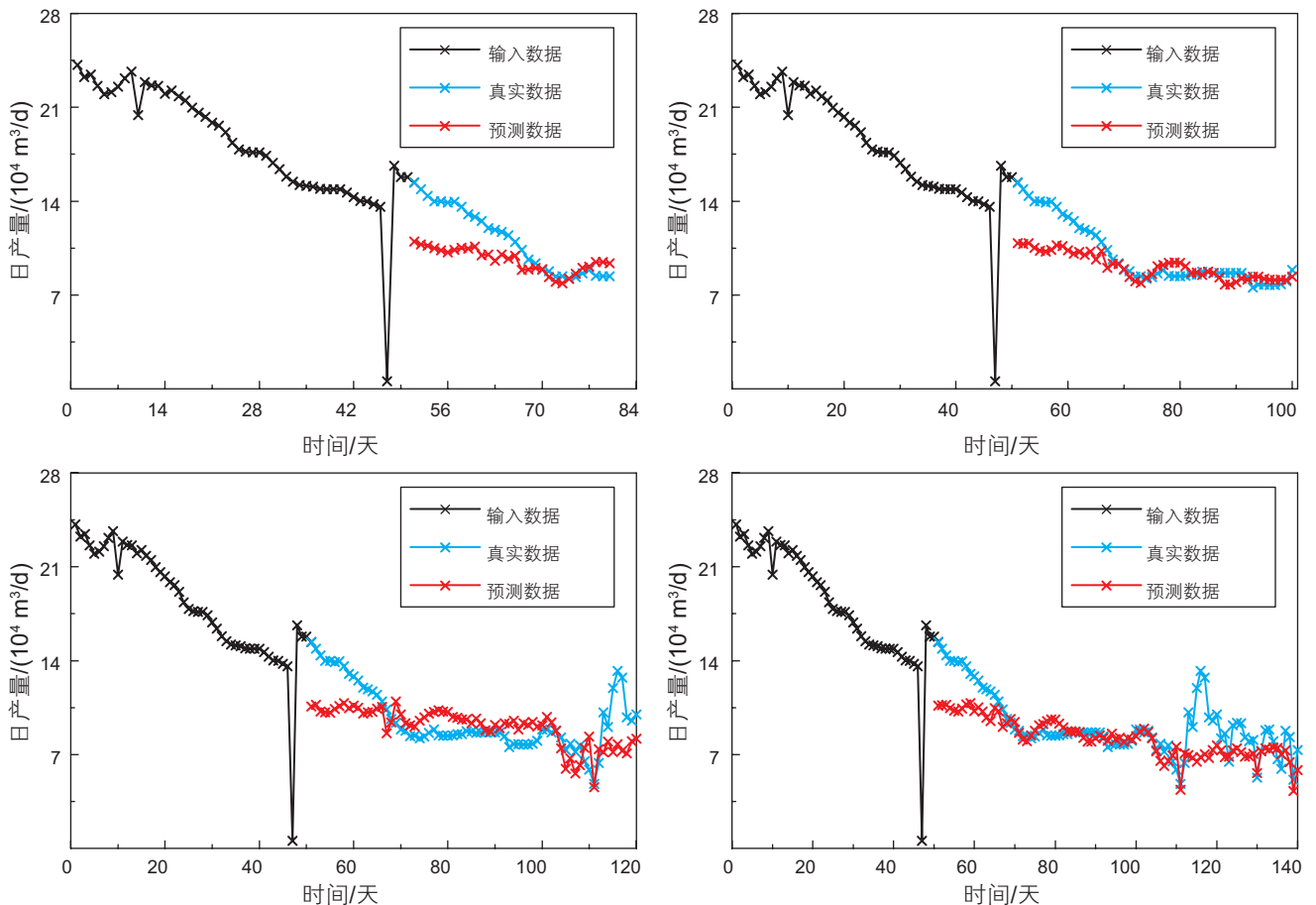


图 8 预测时间步长分别为 30、50、70 和 90 天的 4 井产能预测结果

Fig. 8 Well-4 productivity prediction results for forecast time steps of 30, 50, 70 and 90 days, respectively

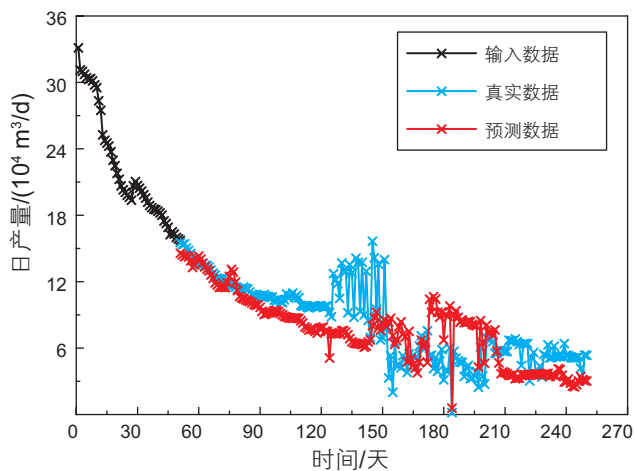


图 9 预测时间步长 200 的 2 井产能预测结果

Fig. 9 Well-2 productivity prediction results for forecast time steps of 200

3 结论

基于 LSTM 和 DNN 神经网络, 本文建立了拟合函数—神经网络协同的动态产能预测模型。模型能够有

效捕捉产能曲线在时间维度和地质及工程维度上的非线性规律。针对某页岩区块的筛选后的 12 口井, 基于 k 折交叉验证分析了神经网络模型参数、静态产能控制参数、输入时间步长和预测时间步长对模型预测精度的影响。结论如下: 1) 通过 LSTM 和 DNN 神经网络的耦合, 可实现多个时序—非时序产能控制参数对后期产能曲线的联合约束; 2) 在数据预处理过程中通过拟合函数对现场数据的规律性判断, 预筛选符合产能递减趋势的样本作为模型训练集, 以间接加入含产能递减规律的弱物理约束, 可在不增加样本规模的条件下能够有效提高模型训练的精度; 3) 基于实际工况下单日生产时间与产量的强相关性, 于神经网络模型内部加入产能等比例缩放的强物理约束, 能够有效提高模型的预测精度和局部稳定性; 4) 激活函数、梯度优化函数分别为 Sigmoid 函数和 Nadam 函数时模型预测误差值为 2.0001, 预测效果最好; 5) 预测时间步长的增加并没有明显降低模型预测的精度, LSTM 神经网络能够有效捕捉长期的时间序列信息。综上, 本文模型考虑了拟合模型的特征筛选, 结合单日生产时间

的强物理约束, 并针对现场产能数据进行训练和预测, 能够较好适应不同特征的产能曲线样本, 并获取对应的产能递减规律, 可以一定程度上辅助页岩气井未来产能曲线预测。

参考文献

- [1] 郎兆新, 张丽华, 程林松. 压裂水平井产能研究[J]. 石油大学学报(自然科学版), 1994, 018(002): 43–46. [LANG Z X, ZHANG L H, CHENG L S. Study on the productivity of fractured horizontal wells[J]. Journal of Petroleum University (Natural Science Edition), 1994, 018(002): 43–46.]
- [2] 宁正福, 韩树刚, 程林松, 等. 低渗透油气藏压裂水平井产能计算方法[J]. 石油学报, 2002, 23(2): 68–71. [NING Z F, HAN S G, CHENG L S, et al. A method for calculating the productivity of fractured horizontal wells in low-permeability reservoirs[J]. Journal of Petroleum, 2002, 23(2): 68–71.]
- [3] CHEN Z, LIAO X, HUANG C, et al. Productivity estimations for vertically fractured wells with asymmetrical multiple fractures[J]. Journal of natural gas science and Engineering, 2014, 21: 1048–1060.
- [4] LUO W, TANG C, WANG X. Pressure transient analysis of a horizontal well intercepted by multiple non-planar vertical fractures[J]. Journal of Petroleum Science & Engineering, 2014, 124: 232–242.
- [5] 王本成, 贾永禄, 李友全, 等. 多段压裂水平井试井模型求解新方法[J]. 石油学报, 2013, 34(6): 1150–1156. [WANG B C, JIA Y L, LI Y Q, et al. A new method for solving test well models of multi-stage fractured horizontal wells[J]. Journal of Petroleum, 2013, 34(6): 1150–1156.]
- [6] KARIMI-FARD M, DURLOFSKY L J, AZIZ K. An efficient discrete-fracture model applicable for general-purpose reservoir simulators[J]. SPE Journal, 2004, 9(2): 227–236.
- [7] LANGSRUD O. Simulation of two-phase flow by finite element methods[C]//All Days. Los Angeles, California: SPE, 1976: SPE-5725-MS.
- [8] DING M, XU S, LI X, et al. Pressure transient analysis of multiple fractured horizontal well in composite shale gas reservoirs by boundary element method[J]. Journal of Petroleum Science & Engineering, 2018, 162: 84–101.
- [9] ARPS J J. Analysis of decline curves[J]. Transactions of the AIME, 1945, 160(01): 228–247.
- [10] 翁文波. 预测论基础[M]. 北京: 石油工业出版社, 1984. [WENG W B. Fundamentals of prediction theory [M]. Beijing: Petroleum Industry Press, 1984.]
- [11] 朱春梦, 蓝兴英. 基于粒子群优化算法的化工稳态流程模拟参数优化[J]. 石油科学通报, 2022, 7(01): 50–60. [ZHU C M, LAN X Y. Optimization of chemical steady-state process simulation parameters based on particle swarm optimization algorithm[J]. Petroleum Science Bulletin, 2022, 7(01): 50–60.]
- [12] 罗刚, 肖立志, 史燕青, 等. 基于机器学习的致密储层流体识别方法研究[J]. 石油科学通报, 2022, 7(01): 24–33. [LUO G, XIAO L Z, SHI Y Q, et al. Machine learning-based fluid identification method for dense reservoirs[J]. Petroleum Science Bulletin, 2022, 7(01): 24–33.]
- [13] 宋先知, 姚学喆, 李根生, 等. 基于LSTM-BP神经网络的地层孔隙压力计算方法[J]. 石油科学通报, 2022, 7(01): 12–23. [SONG X Z, YAO X Z, LI G S, et al. LSTM-BP neural network-based pore pressure calculation method[J]. Petroleum Science Bulletin, 2022, 7(01): 12–23.]
- [14] 程悦菲, 杨洋. 人工智能在油井产量预测上的应用综述[J]. 化工设计通讯, 2021, 47(01): 125–126. [CHENG Y F, YANG Y. A review of the application of artificial intelligence to oil well production prediction[J]. Chemical Design Letters, 2021, 47(01): 125–126.]
- [15] 黄家宸. 基于机器学习的水平井产量预测建模与应用研究[D]. 北京: 中国地质大学(北京), 2020. [HUANG J C. Research on modeling and application of horizontal well production prediction based on machine learning [D]. Beijing: China University of Geosciences (Beijing), 2020.]
- [16] 金宝强, 童凯军, 孙红杰, 等. 一种时变耦合模型在油井产量预测中的应用[J]. 科技导报, 2010, 28(17): 72–76. [JIN B Q, TONG K J, SUN H J, et al. Application of a time-varying coupled model to oil well production prediction[J]. Science and Technology Herald, 2010, 28(17): 72–76.]
- [17] GUO Z, CHEN C, GAO G, et al. Integration of support vector regression with distributed gauss-newton optimization method and its applications to the uncertainty assessment of unconventional assets[J]. SPE Reservoir Evaluation & Engineering, 2018, 21: 1007–1026.
- [18] 徐慧. 粒子群优化算法改进及其在煤层气产能预测中的应用研究[D]. 中国矿业大学, 2013. [XU H. Particle swarm optimization algorithm improvement and its application in coalbed methane productivity prediction[D]. China University of Mining and Technology, 2013.]
- [19] 田冷, 何顺利, 顾岱鸿, 等. 改进BP神经网络模型在长庆气田产能预测中的应用[J]. 石油天然气学报, 2008, 30(05): 106–109+380. [TIAN L, HE S L, GU D H, et al. Application of improved BP neural network model in productivity prediction of Changqing

- gas field[J]. *Journal of Petroleum and Natural Gas*, 2008, 30(05): 106–109+380.]
- [20] CLAR F H, MONACO A. Data-driven approach to optimize stimulation design in eagle ford formation[C]//*Proceedings of the 7th Unconventional Resources Technology Conference*. Denver, Colorado, USA: American Association of Petroleum Geologists, 2019.
- [21] 周于皓, 刘慧卿, 祁鹏, 等. 基于循环神经网络的缝洞型油藏油井产量预测[J]. *计算物理*, 2018, 35(06): 668–674. [ZHOU Y H, LIU H Q, QI P, et al. Well production prediction based on recurrent neural network for slit-hole type reservoirs[J]. *Computational Physics*, 2018, 35(06): 668–674.]
- [22] SAGHEER A, KOTB M. Time series forecasting of petroleum production using deep lstm recurrent networks[J]. *Neurocomputing*, 2019, 323: 203–213.
- [23] 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(02): 77–81+131. [GU J W, ZHOU M, LI Z T, et al. A data mining-based method for oil well production prediction with long and short-term memory network model[J]. *Special Oil and Gas Reservoirs*, 2019, 26(02): 77–81+131.]
- [24] ARPS J J. Estimation of primary oil reserves[J]. *Transactions of the AIME*, 1956, 207(01): 182–191.
- [25] 陈元千, 郝明强. Arps 递减微分方程的推导及应用[J]. *断块油气田*, 2014, 21(01): 57–58. [CHEN Y Q, HAO, M Q. Derivation and application of the Arps decreasing differential equation[J]. *Broken Block Oil and Gas Field*, 2014, 21(01): 57–58.]
- [26] 谢军, 李平, 候康柱. 利用二元回归曲线拟合法和筛选法分析产能递减规律[J]. *新疆石油天然气*, 2005(02): 37–40+44–4. [XIE J, LI P, HOU K Z. Analysis of the law of diminishing productivity using binary regression curve fitting method and screening method[J]. *Xinjiang Oil and Gas*, 2005(02): 37–40+44–4.]
- [27] KLAUS G, K. S R, JAN K, et al. LSTM: a search space odyssey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(10): 2222–2232.
- [28] 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测[J]. *北京航空航天大学学报*, 2018, 44(04): 772–784. [WANG X, WU J, LIU C, et al. Fault time series prediction based on LSTM recurrent neural network[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2018, 44(04): 772–784.]
- [29] 陈云天. 基于机器学习的测井曲线补全与生成研究[D]. 北京: 北京大学, 2020. [CHEN, Y T. Research on Completing and Generating Logging Curves Based on Machine Learning [D]. Beijing: Peking University, 2020.]
- [30] ZENG B, LIU J, SONG Y, et al. Optimization of fracturing parameters in changning shale based on a new comprehensive multi-factor analysis method[C]. *American Society of Mechanical Engineers Digital Collection*, 2021.
- [31] ALISTARH D, GRUBIC D, LI J, et al. QSGD: communication-efficient sgd via gradient quantization and encoding[C]//*Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017.
- [32] MUKKAMALA M C, HEIN M. Variants of rmsprop and adagrad with logarithmic regret bounds[C]//*International Conference on Machine Learning*. PMLR, 2017: 2545–2553.
- [33] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. *Journal of Machine Learning Research*, 2011, 12(61): 2121–2159.
- [34] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//*Proceedings of the 30th International Conference on International Conference on Machine Learning – Volume 28*. Atlanta, GA, USA: JMLR. org, 2013: III–1139–III–1147.

(编辑 马桂霞)