

# 基于模态分解与LSTM注意表征的 测井曲线重构研究

刘 梦<sup>1,2</sup>, 韩 建<sup>1,2\*</sup>, 曹志民<sup>1,2</sup>, 刘兴斌<sup>1</sup>

(1. 东北石油大学物理与电子工程学院, 黑龙江大庆 163319;

2. 东北石油大学黑龙江省高校共建测试计量技术及仪器仪表研发中心, 黑龙江大庆 163319)

**摘 要:** 测井曲线记录着地层物理性质随深度变化的幅值范围,是测井与地震资料之间的纽带,对储层岩性分析与识别和后续的油气勘探工程十分重要。然而,在实际的测井过程中仪器故障等原因会造成测井曲线缺失的问题,重新测井不仅价格昂贵而且难以实现。针对地质勘探时测井数据时常缺失的问题,本文提出了一种LSTM(Long Short-Term Memory)注意力表征的测井曲线重构方法。同时,对原始测井信号进行两种模态分解,计算分解后得到模态分量与原始信号之间的相关性,去除冗余分量,实现对缺失的测井曲线高效、高精度的人工补全。将该方法用于声波(Acoustic, AC)与密度(Density, DEN)曲线重构实验,并将实验结果与LSTM网络和BP(Back Propagation)神经网络预测的结果进行对比分析。结果表明,LSTM-Attention模型有着更为优异的预测效果,重构后的AC和DEN与原始曲线之间的相关性分别达到了86.8%和74.8%,高于传统LSTM和BP神经网络预测方法。在去除冗余的信号分量后,相关系数分别提高了1.4%和4.0%。同时,本文所提方法预测出的测井曲线具有最低的预测误差。因此,基于LSTM注意表征的网络结构对测井曲线重构具有较好的预测精度。

**关键词:** 长短期记忆人工神经网络;注意力机制;测井曲线重构;VMD;EMD

**基金项目:** 国家自然科学基金(No.52174021)

**中图分类号:** TN911.7;TP391

**文献标识码:** A

**文章编号:** 0372-2112(2024)04-1399-12

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20220887

## Logging Curves Reconstruction Based on Mode Decomposition and LSTM-Attention Model

LIU Meng<sup>1,2</sup>, HAN Jian<sup>1,2\*</sup>, CAO Zhi-min<sup>1,2</sup>, LIU Xing-bin<sup>1</sup>

(1. School of Physics and Electronic Engineering, Northeast Petroleum University, Daqing, Heilongjiang 163319, China;

2. Research and Development Center for Testing and Measurement Technology and Instrumentation, Heilongjiang Province University, Northeast Petroleum University, Daqing, Heilongjiang 163319, China)

**Abstract:** The well logging curve records the amplitude range of geophysical properties changing with depth and is the bond between well log and seismic data. It is also significant for reservoir lithology analysis and subsequent oil and gas exploration projects. However, instrument failure and other reasons will cause well-logging curves to be missing in the actual logging process. Re-logging is not only expensive but also difficult to achieve. Aiming at the problem that logging data is often missing during geological exploration, this paper proposes a logging curve reconstruction method based on the LSTM (Long Short-Term Memory)-attention model. At the same time, EMD-VMD (Empirical Mode Decomposition-Variational Mode Decomposition) decomposition is performed on the original logging signal and then the correlation between the components and the original curve is calculated. Some excrescent components are deleted to promote efficient and high-precision manual completion. This proposed method is applied to reconstruct missing logs acoustic (AC) and density (DEN), and the prediction results are compared with those predicted by LSTM and BP (Back Propagation) neural network. The results show that the LSTM-attention model has a better prediction performance, and the correlations between predictive and the original curves can reach 86.8% (AC) and 74.8% (DEN), higher than the traditional LSTM and BP neural network. After re-

moving redundant signal components, the correlation coefficients increased by 1.4% (AC) and 4.0% (DEN). At the same time, the logging curve predicted by the proposed method has the lowest prediction error. Therefore, the representation learning based on LSTM with an attention mechanism has better prediction accuracy for well-logging curve reconstruction.

**Key words:** long short-term memory; attention; log curves reconstruction; VMD; EMD

**Foundation Item(s):** National Natural Science Foundation of China (No.52174021)

## 1 引言

测井数据沿着井眼记录井的地球物理性质。利用测井技术,地球物理学家和地质学家可以获得各种重要的储层地质参数,如孔隙度、渗透率、油水饱和度、岩性、沉积微面等<sup>[1]</sup>。测井曲线各项参数的观测对于准确建立地质模型至关重要,其可以描述储层的流动特性,改进储层评价,提高钻井效率,为工程生产提供实时策略<sup>[2]</sup>。

受钻井条件差、仪器故障、测井条件差异、存储不当导致的数据丢失、勘探成本等因素的影响,不可避免地会出现测井数据不完整或测井曲线完全缺失的现象<sup>[3-5]</sup>。因此,地球物理学家和地质学家准确估算并提出缺失测井曲线预测方案,对提高油气资源的开发效率至关重要。一般处理预测问题常用的方法有经验模型方法<sup>[3]</sup>、多元线性回归方法<sup>[4]</sup>、传统的机器学习方法<sup>[5-8]</sup>、分形重构方法<sup>[9]</sup>、模糊推理方法<sup>[10]</sup>、时间序列模型<sup>[11]</sup>等。上述研究揭示了不同时序曲线之间的非线性关系。对于测井曲线重构的研究,传统主要采用数学思路,例如克里金插值法<sup>[12]</sup>、物理模型反演法<sup>[13,14]</sup>等。

随着人工智能技术的高速发展,其被逐渐引入了测井曲线的预测工作中。Rolon 等人<sup>[15]</sup>建立了神经网络模型,来对缺失的电阻率曲线进行复原;杨志力等人<sup>[16]</sup>将 BP 神经网络技术运用在声波测井曲线重构中;张东晓等人<sup>[17]</sup>基于长短期记忆(Long Short-Term Memory, LSTM)网络来构建补偿声波和密度曲线;王俊等人<sup>[18]</sup>利用门控循环单元(Gate Recurrent Unit, GRU)寻找测井数据之间的非线性映射关系,该方法相对于 LSTM 具有更快的收敛速度,具有较高的精度和鲁棒性。

测井数据属于离散的时序数据,井下储层的物理属性具有一定的连续性,虽然 LSTM 可以很好地处理序列数据的上下文信息,但是随着井位置的改变,不同井之间的地下地质物理属性差异较大,利用相邻井来预测本井缺失的测井曲线难度很大。注意力机制可以使网络更加关注对缺失曲线有利的特征信息,赋予重要特征更高的权重,进一步提高预测的准确率,很好地弥补了 LSTM 虽可以关注上下文信息但没有着重关注点的缺点。所以,本文在 LSTM 网络的基础上构建了一种 LSTM-Attention 预测模型,对缺失的井下测井数据进行重构。此外,为深度挖掘测井曲线之间的隐藏特征,提

升输入信号的抗噪能力,本文还对测井信号进行了模态分解。为减少输入信号冗余性,对分解后得到的模态分量进行相关性分析,去除与原始信号相关性较低的信号分量,使 LSTM-Attention 模型的输入精简有效,此举进一步提高了模型的预测效率和准确率。

## 2 基本理论

经验模态分解(Empirical Mode Decomposition, EMD)可以根据自身的时间尺度特征对非线性测井信号分解,在保持原始测井信号良好特性的同时,又可以挖掘出信号的隐藏信息<sup>[19]</sup>;变分模态分解(Variational Mode Decomposition, VMD)采用非递归的处理策略分解测井曲线信号,对非线性复杂的测井曲线具有较好的分解精度和抗噪干扰能力。下面介绍 EMD 与 VMD 的具体算法过程。

### 2.1 EMD

EMD 可以从任何复杂的信号中逐步提取不同尺度的波形或趋势,挖掘出时序数据的隐藏关系。它可以信号分解为若干个固有本征模态函数(Intrinsic Mode Function, IMF)和一个残余项,其步骤<sup>[20,21]</sup>如下:

(1) 根据时序信号  $x(t)$  的局部极大值和局部极小值找出信号的上下包络线  $\max_x(t)$  和  $\min_x(t)$ ;

(2) 求出上下包络的均值,得出均值包络线  $m_1(t)$ :

$$m_1(t) = \frac{\min_x(t) + \max_x(t)}{2} \quad (1)$$

(3) 信号  $x(t)$  减去均值包络得到中间信号  $c_1(t)$ :

$$c_1(t) = x(t) - m_1(t) \quad (2)$$

(4) 以  $c_1(t)$  代替原始时序信号  $x(t)$ , 并重复步骤(1)~(3), 直至信号  $c_1(t)$  满足两个条件: 一是局部极值点的数目与过零点的数目相同或仅相差一个; 二是该分量的局部极大值的上包络线与局部极小值的下包络线在任意时刻的均值为零。此时的  $c_1(t)$  便为第一个固有模态函数(IMF1)。

(5) 从原始时序信号  $x(t)$  中分离 IMF1, 剩余的残差序列  $r_1(t)$  为

$$r_1(t) = x(t) - \text{IMF}_1 \quad (3)$$

(6) 将  $r_1(t)$  视为新的原始信号, 重复步骤(1)~(5), 直至残差信号  $r_n(t)$  为单调函数或者设定的阈值;

(7) 最终得到了符合条件的一系列模态分量  $c_i$  和

$r_n(t)$ , 此时原始时序信号  $x(t)$  被分解为

$$x(t) = \sum_{i=1}^n c_i + r_n(t) \quad (4)$$

### 2.2 VMD

VMD 根据迭代来找到变分模型的最优解, 从而确定信号的本征模态函数 (Intrinsic Mode Functions, IMFs) 分量和对应的具有不同中心频率的有效带宽, 可以实现信号频谱的划分和各本征模态函数分量的有效分离, 主要计算过程<sup>[22-26]</sup>如下。

(1) 采用 Hilbert 变换计算出每个模态函数  $u_k(t)$  的解析信号, 以获得其单边频谱。

(2) 通过各个模态解析信号与所对应的中心频率  $e^{-j\omega_k t}$  项混合, 将各模态的频谱调制到基频带。

(3) 依据高斯平滑度和梯度平方准则对信号进行解调, 进而得到各个分解模态的带宽。分解后的各模态量都为调幅—调频信号, 其变分约束模型为

$$\min \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (5)$$

$$\text{s.t.} \sum_{k=1}^K u_k = f \quad (6)$$

(4) 求变分约束模型的最优解, 将上述的约束性问题转换为非约束性问题。得到的拉格朗日函数为

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \alpha \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - [u_k(t)] \right\|_2^2 \\ &+ \left[ \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right] \end{aligned} \quad (7)$$

其中,  $\alpha$  为二次惩罚因子;  $\lambda$  为拉格朗日乘数。

在非约束问题上, 采用交替方向乘子法, 通过交替更新, 最终可以获得  $u_k^{n+1}$ ,  $\omega_k^{n+1}$  和  $\lambda_k^{n+1}(t)$  ( $n$  为迭代次数), 即式(5)中的约束变分模型的最优解:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{u}_i(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (8)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (9)$$

在式(8)和式(9)中,  $\hat{u}_k^{n+1}(\omega)$  为当前余量的维纳滤波;  $\omega_k^{n+1}$  为当前模态函数的中心频率。

迭代停止的条件为

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|^2}{\hat{u}_k^n} < \zeta \quad (10)$$

其中,  $\zeta$  为给定的判定精度。

### 2.3 LSTM

LSTM 是一种特殊的 RNN (Recurrent Neural Network)。它的重复网络功能由 3 个门结构来实现, 整体的网络结构可以很好地模仿神经元传递信息的模式, 储存更长的时间序列信息<sup>[27,28]</sup>。LSTM 的结构如图 1 所示。

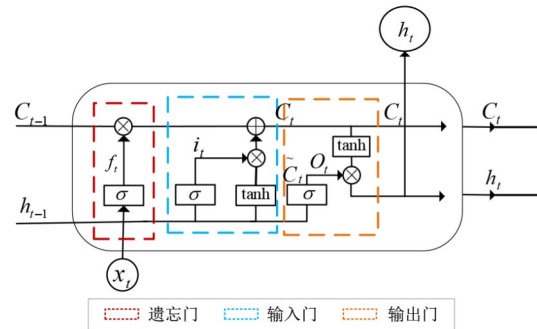


图 1 LSTM 的结构图

(1) 遗忘门: 决定当前步骤忘记了什么信息,  $f(t)$  等于 0 时, 遗忘所有信息。如式(11)所示:

$$f(t) = \sigma[W_f(h_{t-1}, x_t) + b_f] \quad (11)$$

(2) 输入门: 输出  $i_t$  决定了添加到单元状态的新信息。如式(12)所示:

$$i_t = \sigma[W_i(h_{t-1}, x_t) + b_i] \quad (12)$$

输入门层中包含的 tanh 层, 创建了可以添加到单元状态的新候选值  $\tilde{C}_t$ , 如式(13)所示:

$$\tilde{C}_t = \tanh[W_c(h_{t-1}, x_t) + b_c] \quad (13)$$

此时, 单元状态被更新,  $C_t$  由两部分组成, 如式(14)所示:

$$C_t = C_{t-1}f(t) + i_t\tilde{C}_t \quad (14)$$

(3) 输出门: 决定了更新后的单元状态生成 LSTM 的输出值, 输出值为

$$h(t) = \sigma[W_o(h_{t-1}, x_t) + b_o] \quad (15)$$

在式(11)~(15)中,  $\sigma$  和  $\tanh$  为激活函数;  $x_t$  为某时刻的输入样本;  $W$  权重;  $b$  为偏置项。

## 3 测井曲线的预测模型设计

### 3.1 LSTM-Attention 预测模型

注意力机制通过识别输入的高相关性提高模型的学习能力, 已经被成功用于多个领域, 例如翻译、图像分类、语音处理等<sup>[29,30]</sup>。在做时间序列数据预测时, 注意力机制通过综合分析输入的特征, 并对特征的相关性进行权重分配, 可以使预测模型缩小对输入数据的

关注范围,模型在训练过程中可以更加密切地关注关键信息,增强自身的预测性能.

注意力机制通过对所有的输入变量加权平均,得到了下一层的输入变量. 对于一个长度为  $L$  的输入数据  $V_i$ ,对其进行加权注意的通用表达式为

$$\text{Attention}(Q, S) = \sum_{i=1}^L \text{softmax}(Q_i, K_i) \cdot V_i \quad (16)$$

其中,  $Q$  为概率矩阵;  $K$  为键; 通过  $\text{softmax}()$  归一化后, 加权求和. 计算过程如图 2 所示.

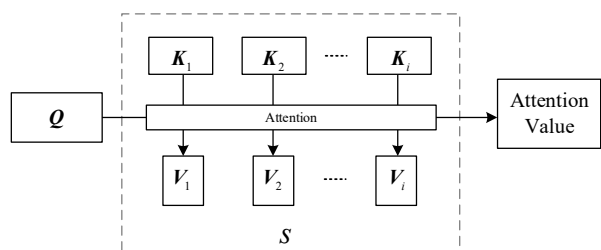


图2 Attention Value 计算方法

本文的目标是使用相邻井已有的测井曲线来重构本井内完全缺失的测井曲线, 可以将测井数据看作随深度而变化的时间序列. 使用 LSTM 网络作为基础的预测模型, 提取测井数据随深度而变化的特征. 然后, 在 LSTM 网络层后加入注意力机制层, 赋予输出值不同的权重, 最后加入全连接层. LSTM-Attention 预测模型构建如图 3 所示.

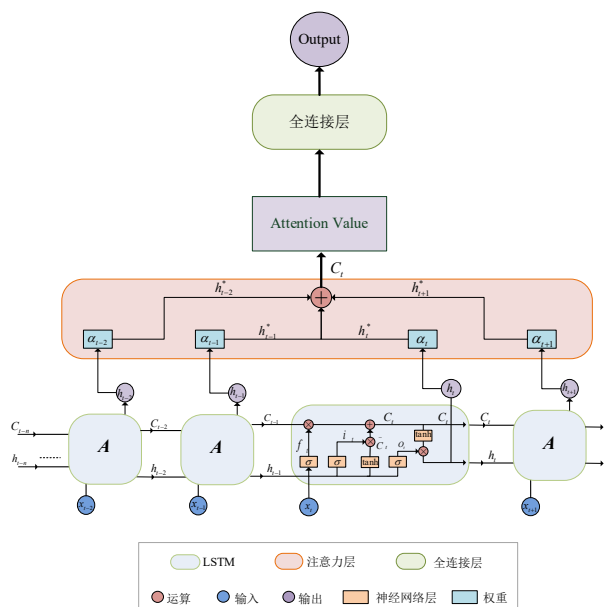


图3 LSTM-Attention 预测模型

在有注意力机制的 LSTM 网络中, 注意力机制层根据 LSTM 的最后一个单元状态与当前的状态校准, 计算出当前状态与中间隐藏状态之间的相关性. 在学习过

程中, 通过赋予不同的权重, 凸显出相关的信息, 抑制不相关的信息, 从而可以提高预测的准确率和效率.

由式(11)和式(15)可知, 长度为  $L$  在  $t$  时刻的输入数据  $x_t$  经过 LSTM 的输出值为  $h_t$ , 即

$$h_t = \text{LSTM}(x_t) \quad (17)$$

计算 Attention 层隐藏状态中的特征向量  $h'_t$  与  $h_t$  之间的相关性, 即

$$P_t = \text{cor}([h_t, h'_t]) \quad (18)$$

对不同时刻的  $h'_t$  进行权重配置, 高相关的  $h'_t$  通过概率矩阵赋予更高的权重. 其权重概率矩阵  $Q_i$  如式(19)所示:

$$Q_i = \text{softmax}(\sigma(P_t, H_t) + b) \quad (19)$$

其中,  $H_t$  为隐藏层;  $\sigma$  为单层感知机;  $b$  为偏置向量. 可以得到注意力模型的输出  $C_i$ :

$$C_i = \sum_{i=1}^L Q_i \cdot h'_i \quad (20)$$

最后, 将  $C_i$  输入到一个全连接层中, 可以得到最终的预测值  $Y$  为

$$Y = WC_i + b \quad (21)$$

其中,  $W$  为全连接层的权重,  $b$  为偏置.

### 3.2 油井测井曲线预测模型搭建

如图 4 所示, 实验框架的搭建共分为两步: 数据选择与模型参数设置.

#### 3.2.1 数据选择与处理

选择中国大庆油田古区的井 Gu92 和 Gu93 作为训练井和测试井, 这两口井均位于松辽盆地的中央拗陷区. 训练井 Gu92 的起止深度为井下 1 215~2 229.5 m, 每隔 0.125 m 有一个测井点, 共 8 117 个测井点, 全长 1 014.5 m. 测试井 Gu93 的起止深度为井下 1 957~2 302 m, 每隔 0.125 m 有一个测井点, 共 2 760 个测井点, 全长 345 m. 两口井同一深度下的部分精细岩层对比如图 5 所示.

由图 5 可以看出, 训练井 Gu92 的储层岩性主要包含泥岩、粉砂质泥岩、泥质粉砂岩等常见的岩性. 而测试井 Gu93 的储层岩性与训练井差异较大, 除了包含常见的岩性外, 还包含了多种不同类型的岩性层. 这表明尽管两口井处在同一区域, 但是井下的物理属性可能差异较大, 这为后续的测井曲线预测增加了一定难度.

两口油井均包含 7 条常规曲线: 声波 (ACoustic, AC)、井径 (CALiper, CAL)、自然伽马 (Gamma Ray, GR)、自然电位 (Spontaneous Potential, SP)、深侧向电阻率 (LateroLog Deep, LLD)、浅侧向电阻率 (LateroLog Shallow, LLS)、密度曲线 (DENsity, DEN). 在实际石油勘探工程中, 曲线 AC 与地质的地层密度密切相关, 它对判断储层岩性和计算矿物含量具有重要的物理意

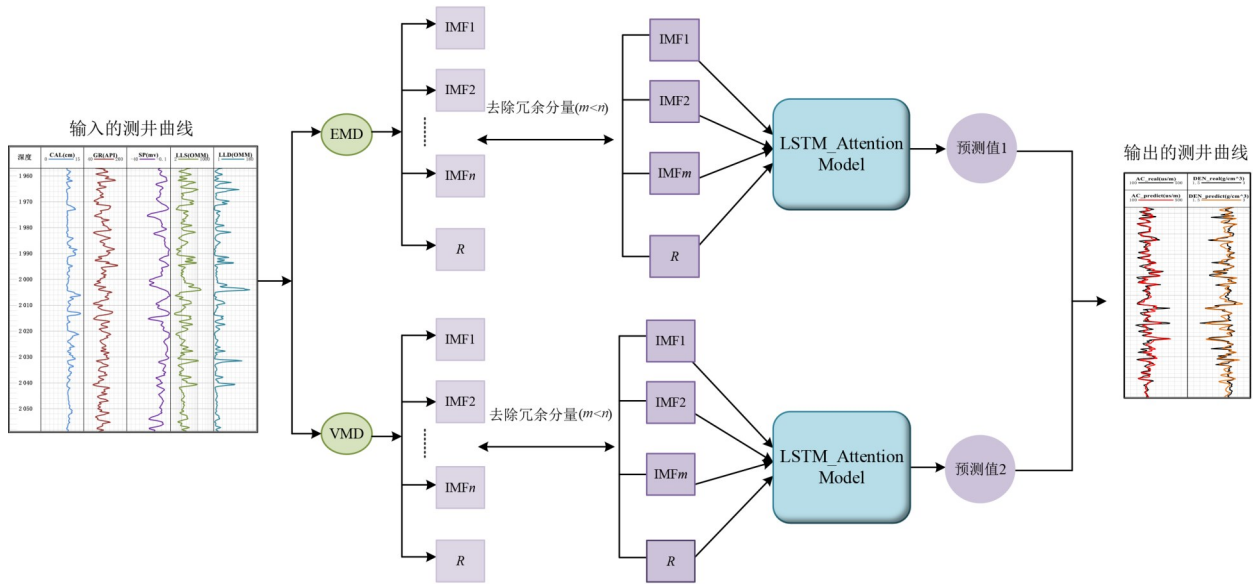


图4 实验整体框架

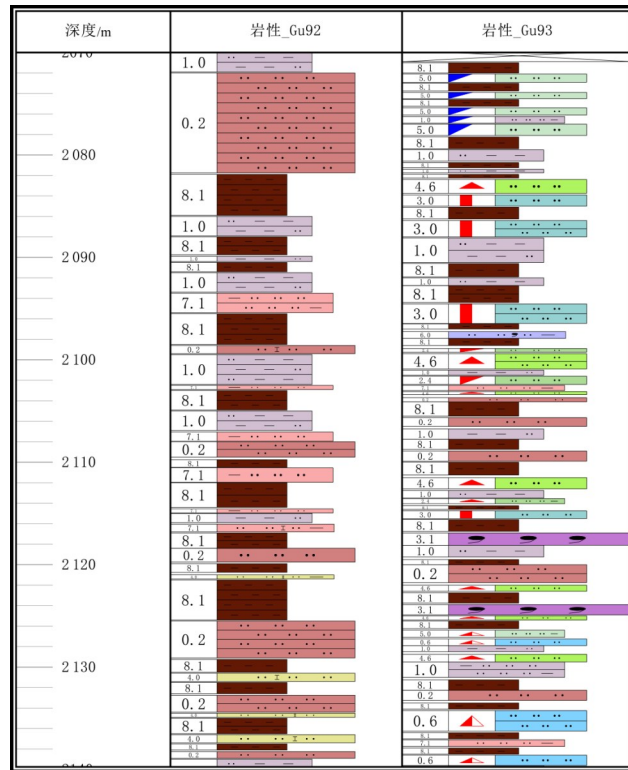


图5 两口实验井的岩层对比

义;利用密度曲线 DEN,可以很好地识别岩性,并确定岩层的孔隙度.此两条曲线的完整性对后续的工程具有实际意义,因此后续实验选择曲线 AC 和 DEN 来进行人工重构.

由于测井仪器运动测量、井下环境的时移改变等,初步获得的测井数据含大量噪声<sup>[31]</sup>,首先需要对测井数据进行去除野值点处理,提高预测精度<sup>[32,33]</sup>.部分

测井数据集及去除野值点后的数据集如图6所示.

为确保每个参数都处于同一分布范围内,避免模型的预测性能受到不同量纲的数据影响,需要对测井数据进行标准化处理,标准化公式如下:

$$y^* = \frac{y - \mu}{\sigma} \tag{22}$$

其中, $\mu$ 和 $\sigma$ 分别为样本数据的均值和标准差; $y$ 为输入

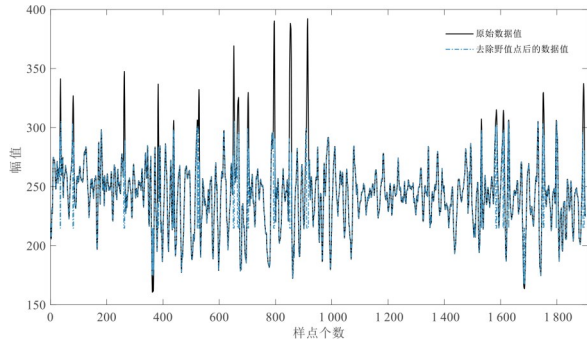


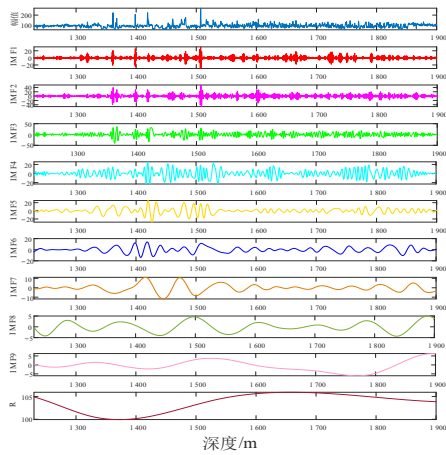
图6 部分测井数据集

数据;  $y^*$  为标准化后的数据. 标准化还具有加速训练过程和防止神经网络梯度爆炸的优点<sup>[34]</sup>.

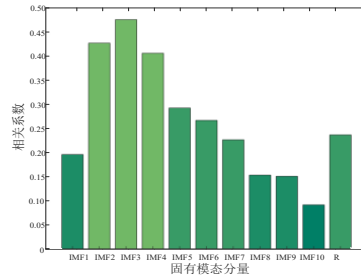
对预处理后的测井数据进行EMD、VMD分解,可以得到不同的IMF分量. 利用皮尔逊系数(Pearson Correlation Coefficient, PCCs)的大小来衡量模式分量与原数据之间的相关性,计算方式为

$$PCCs = \frac{\text{cov}(x_i, x'_i)}{\sigma_{x_i} \cdot \sigma_{x'_i}} \quad (23)$$

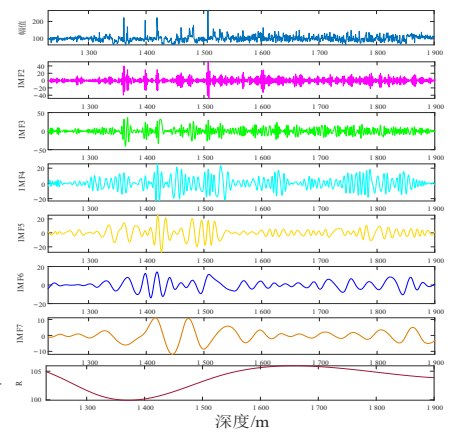
由统计学知识可知,当PCCs小于0.2时,代表两向量之间处于极弱相关或无相关状态. 将PCCs小于0.2的分量去除,减少信息的冗余性,提高模型训练的效率. 输入PCCs更高的IMF分量,提升曲线预测的精度. 以GR曲线为例,对其进行EMD、VMD分解后的相关性分析与冗余信息去除,过程如图7和图8所示.



(a) 对GR曲线进行EMD分解

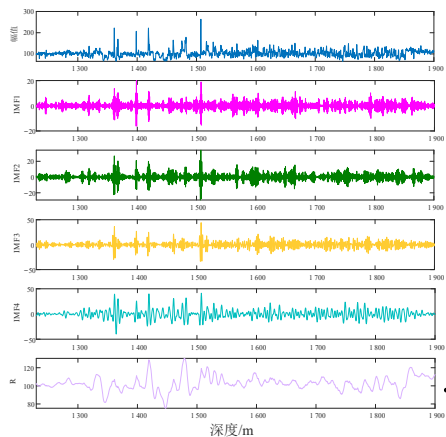


(b) IMFs与原始GR曲线的相关性分析

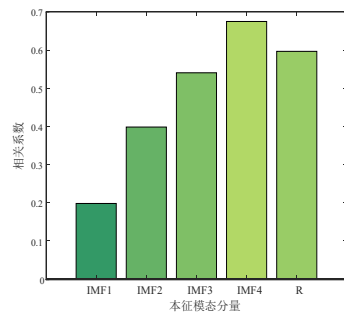


(c) 筛选出相关系数大于0.2的IMF分量

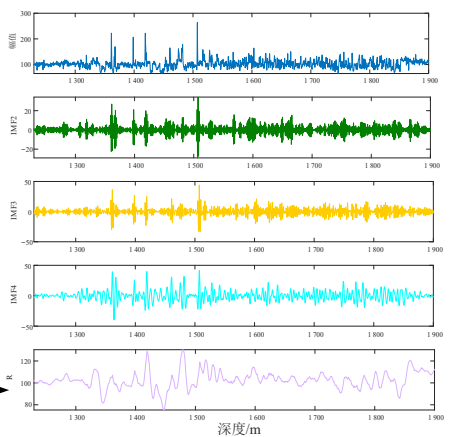
图7 对测井信号EMD分解后并去除冗余分量



(a) 对GR曲线进行VMD分解



(b) IMFs与原始GR曲线的相关性分析



(c) 筛选出相关系数大于0.2的IMF分量

图8 对测井信号VMD分解后并去除冗余分量

### 3.2.2 参数配置

本文使用的实验平台为MATLAB 2020a,图3中的LSTM-Attention模型具体参数的设置如表1所示.

然后对LSTM-Attention模型构架进行参数设置. 依次对批处理大小、训练次数和学习率配置的初始值

为(16,10,0.05),采用贪婪策略搜索最优的参数组. 具体的参数搜索过程如表2所示,加粗数据表示最优结果. 随机选取训练井Gu92中的90%的数据作为训练数据,剩下的10%的数据作为验证数据,来更新模型参数,损失函数的收敛过程如图9所示.

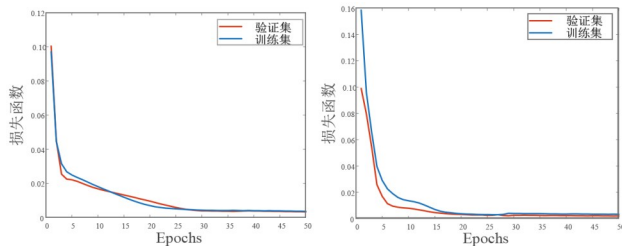
表1 网络模型参数设置

网络层	参数	输出尺寸
输入层	CAL,LLD,LLS,GR,SP	(N,1)
LSTM层	单元个数为100,层数为4	(None,100)
Attention层	单元个数为100,层数为1	(None,100)
Attention输出层	单元个数为100,层数为1	(None,100)
Dense层	单元个数为1	(None,1)

表2 每组参数配置的误差值

Batch_size	Epochs	学习率	MSE/%
16	10	0.050	11.682
32	10	0.050	11.347
64	10	0.050	<b>11.284</b>
128	10	0.050	11.288
64	10	0.050	11.284
64	<b>30</b>	0.050	<b>10.625</b>
64	50	0.050	10.625
64	30	0.050	10.625
64	30	0.010	9.821
64	30	<b>0.005</b>	<b>9.232</b>
64	30	0.001	9.247

图9和表2的结果显示,模型在训练约30个epoch后损失函数已经趋于收敛状态,所以为节省训练时间,最好的参数组合为(64,30,0.005).因此,模型的batch\_size设置为64,训练次数为30,初始学习率为0.005,每轮的训练周期结束后,乘以因子0.5来降低学习率.训练过程中选择均方误差(Mean Squared Error, MSE)作为损失函数,来评价模型的收敛速度,如式(23)所示.此外,Adam作为更新LSTM权重和参数的优化器;为防止预测模型出现梯度爆炸的风险,采用SqGrad梯度下降法来训练深层模型.



(a) 预测AC曲线 (b) 预测DEN曲线

图9 模型损失函数收敛图

为了衡量LSTM-Attention模型的预测效果,采用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)、平均绝对相对误差(Mean Absolute Percentage Error, MAPE)、对称平均绝对百分比误差(Symmetric Mean Absolute Percentage Er-

ror, SMAPE)作为评价指标,如式(24)~(28)所示.此类指标数值越小,测井曲线的重构情况越好.式(23)中的PCCs用来评价真实值与预测值之间的相关性,PCCs越接近1,预测效果越好.

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - x'_i)^2 \quad (24)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - x'_i| \quad (25)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x'_i)^2} \quad (26)$$

$$MAPE = \sum_{i=1}^N \left| \frac{x_i - x'_i}{x_i} \right| \times \frac{100\%}{N} \quad (27)$$

$$SMAPE = \frac{100\%}{N} \times \sum_{i=1}^N \frac{|x_i - x'_i|}{\frac{|x_i| + |x'_i|}{2}} \quad (28)$$

式(23)~(28)中, $x_i$ 为真实值; $x'_i$ 为预测值; $N$ 为样点个数; $cov(\cdot)$ 为协方差; $\sigma$ 为标准差.

### 4 实验与分析

为分别验证LSTM-Attention模型的重构性能和加入模态分量后模型的重构性能,本文共进行两组实验.在两组实验中,两口井内的整条声波时差AC曲线和密度DEN曲线均人为删除,以模拟缺失的测井曲线.训练集为井Gu92,测试集为井Gu93.

实验1:验证LSTM-Attention模型的预测性能.

本实验中,采用原始LSTM模型和BP神经网络模型来对比验证LSTM-Attention模型的预测性能.不同深度下的缺失段AC和DEN的预测结果如图10和图11所示.图10和图11中,黑色曲线为原始真实数据,红色曲线为LSTM-Attention模型重构的结果,其他颜色曲线为LSTM和BP的预测结果,表3和表4中为各类重构定量结果,加粗数据表示最优结果.

如图10和图11所示,LSTM-Attention模型预测出的两条曲线的整体走势均更加贴近真实曲线的走势.在图10波谷和波峰的局部放大图中可以看出,LSTM-Attention模型相比LSTM和BP模型可以更好地捕捉到曲线的极值突变,预测出的曲线更加贴合原始AC曲线走势.表3中给的数据显示,LSTM-Attention模型预测结果与真实AC之间的相关系数最高,达到了86.8%,这比LSTM和BP神经网络的预测结果分别高出了2.4%和8.2%.同时,LSTM-Attention预测结果的4个误差中,除RMSE外均为最低,分别比LSTM的预测结果低了0.109(MAE),0.663(MAPE)、0.22(SPAPE),比BP神经网络的预测结果低了2.295(MAE),4.886(RMSE),0.474(MAPE),0.435(SMAPE).

在对DEN预测的实验中,图11随深度局部放大的

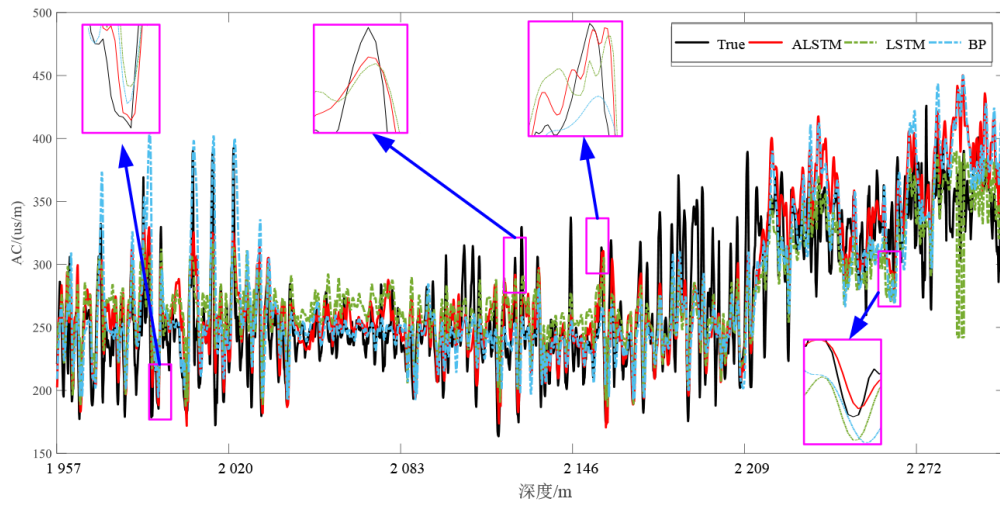


图 10 曲线 AC 在不同模型下的预测结果

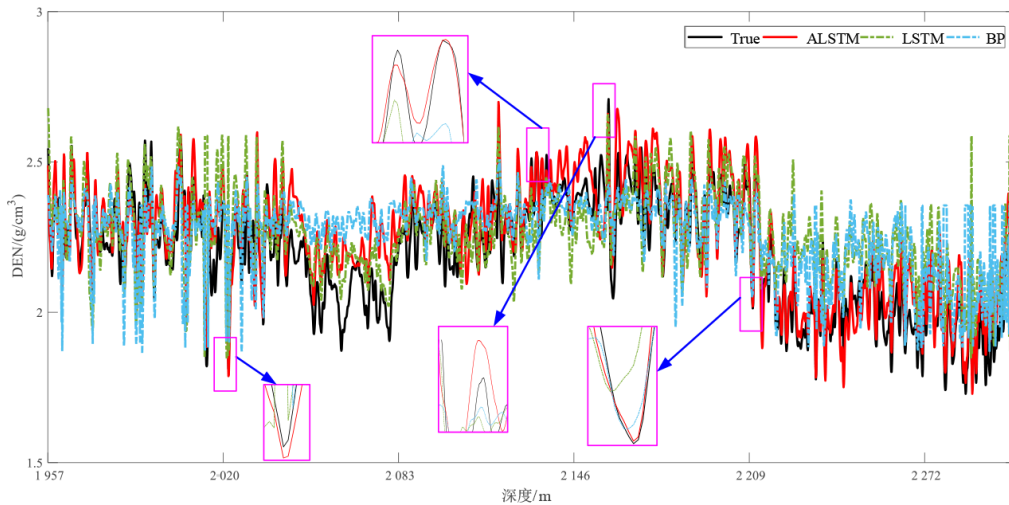


图 11 曲线 DEN 在不同模型下的预测结果

表 3 曲线 AC 的重构定量结果

方法	MAE	RMSE	MAPE	SMAPE	PCCs
LSTM-Attention	21.439	29.452	8.063	7.821	0.868
LSTM-Attention-EMD-VMD	<b>18.756</b>	<b>25.372</b>	<b>6.997</b>	<b>6.874</b>	<b>0.882</b>
LSTM	21.548	27.958	8.726	8.041	0.841
LSTM-EMD-VMD	33.924	40.028	13.261	12.584	0.850
BP	23.734	34.338	8.537	8.256	0.786
BP-EMD-VMD	25.019	32.454	9.574	7.882	0.823

表 4 曲线 DEN 的重构定量结果

方法	MAE	RMSE	MAPE	SMAPE	PCCs
LSTM-Attention	0.215	0.225	7.838	0.078	0.748
LSTM-Attention-EMD-VMD	<b>0.103</b>	0.127	<b>3.398</b>	0.044	<b>0.788</b>
LSTM	0.191	0.185	3.751	<b>0.037</b>	0.697
LSTM-EMD-VMD	0.112	<b>0.125</b>	4.253	0.042	0.714
BP	0.243	0.247	9.286	0.109	0.678
BP-EMD-VMD	0.128	0.126	4.754	0.052	0.694

预测图中,红色曲线相比其他曲线明显更加贴合黑色曲线的突变走势.在表4中,尽管LSTM-Attention模型的预测值与真实值之间的PCCs仅能达到74.8%,但依旧比LSTM和BP神经网络的预测结果的PCCs高,分别高了5.1%和7.0%.这表明添加注意力机制后的LSTM网络可以各以更好地关注重要特征信息,提升预测的准确率.但是,LSTM预测结果各类复原误差值是三个模型中最低的,分别为0.191(MAE),0.185(RMSE),3.751(MAPE),0.037(SMAPE).

实验2:验证模态分量对模型预测性能影响.

在本实验中,我们将去除冗余分量的分解信号共同输入到LSTM-Attention,LSTM,BP模型中,由于篇幅原因,仅展示LSTM-Attention模型添加模态分量后的重构结果.在图12和图13中,黑色曲线为真实值,蓝色曲线为LSTM-Attention模型的重构结果,红色曲线为添加模态分量后的LSTM-Attention模型重构结果.3个模型添加模态分量后

的重构定量结果展示在图 14、表3和表4中。

图 12 和表 3 中,3 组预测模型在输入去除冗余的模态分量后,AC 的预测结果均得到了提升,相关系数分别提升了 1.4%(LSTM-Attention),0.9%(LSTM),3.7%(BP)。其中,LSTM-Attention 的相关系数最高,达到了 88.2%,这表明采用此方法重构的 AC 与原始 AC 处于极度相关状态。图 12 中显示,不同深度下,LSTM-Attention 综合模型重构出的 AC 更加贴合原始 AC 的曲线走势。

表 3 和图 14(b)中 3 组综合模型的重构定量结果显示,在输入模态分量后,LSTM-Attention 模型的预测误

差减少,比初始的 LSTM-Attention 模型的重构结果低了 2.683 (MAE), 4.08 (RMSE), 1.066 (MAPE), 0.94 (SPAPE), 并且 4 个误差为 6 组实验中的最小值,分别为 18.756(MAE), 25.372(RMSE), 6.997(MAPE), 6.874 (MAPE)。所以,EMD-VMD 分解后的模态分量对于提升模型的预测效果具有一定的贡献。

因此,本文所提的 LSTM-Attention 综合模型与 LSTM 和 BP 神经网络模型相比,对曲线 AC 具有较好的预测效果,加入处理后的模态分量后 AC 的预测精度得到进一步提升,重构出的测井曲线与原始曲线之间具

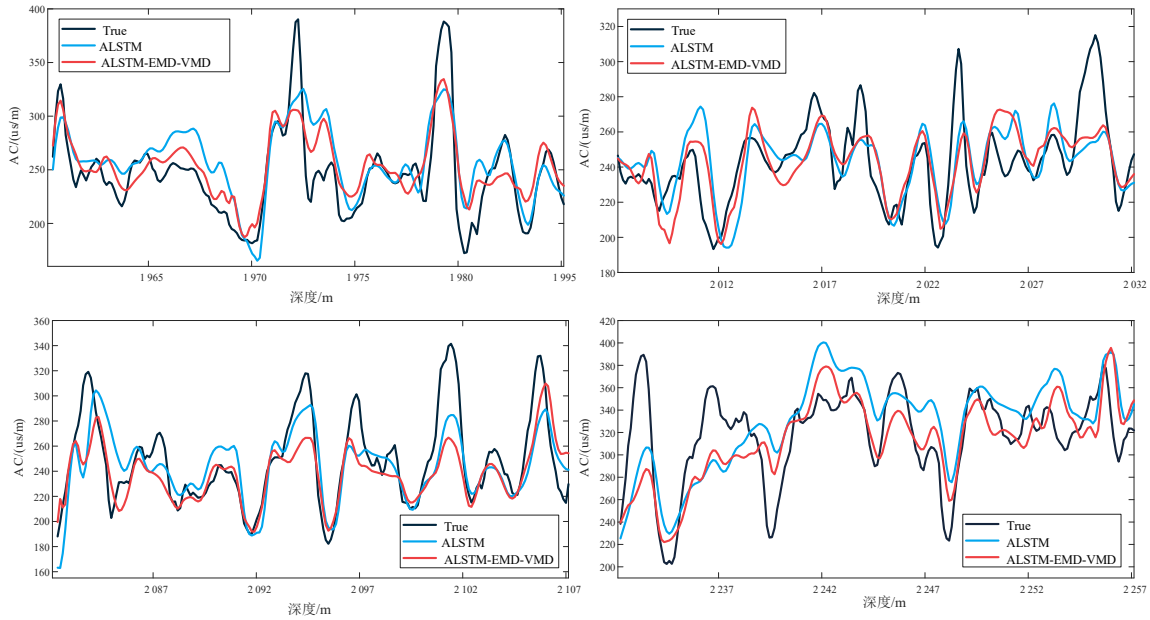


图 12 测试集中部分样点的 AC 预测结果

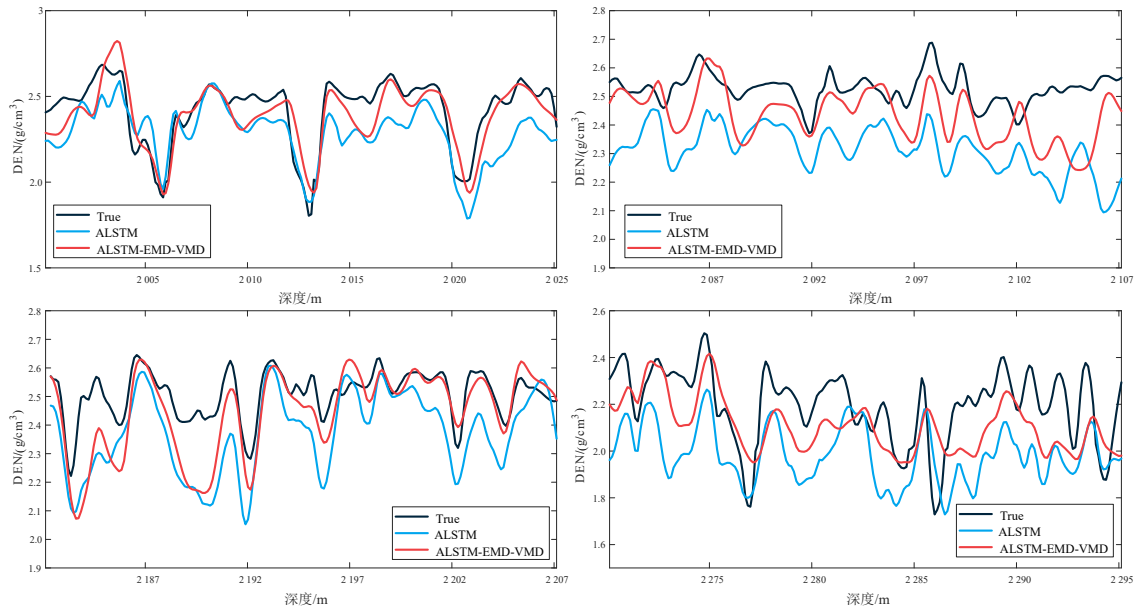
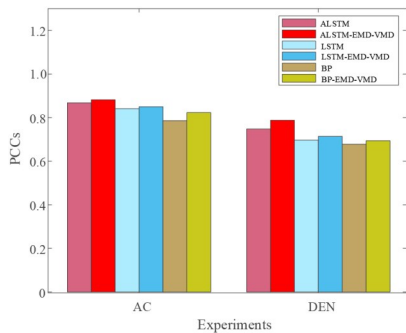


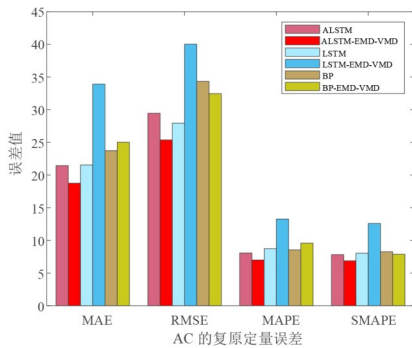
图 13 测试集中部分样点的 DEN 预测结果

有高度相关性和更低的误差。

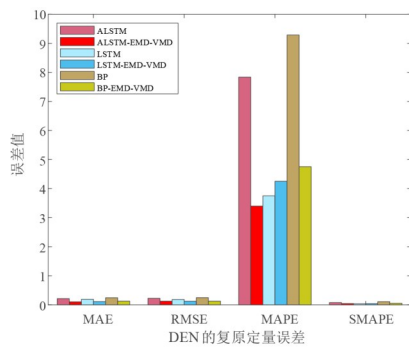
对 DEN 进行重构的实验结果呈现在图 14(a) 和图 14(c) 与表 4 中。LSTM-Attention 综合模型预测的 DEN 与真实 DEN 之间的 PCCs 虽仅能达到 78.8%, 但是却比其他 5 组实验的 PCCs 都要高, 且达到了强相关状态。加入模态分量后, LSTM 和 BP 神经网络模型的 PCCs 也得到了提升, 分别高了 1.7% 和 1.6%。值得注意的是, LSTM-Attention 综合模型的重构结果的 PCCs 比 LSTM-Attention 原始模型高了 4.0%。所以, EMD-VMD 分解得到的模态分量在进行冗余信息的去除后, 可以更好地挖掘出测井曲线之间的非线性关系, 提高模型的预测效率与准确率。在图 13 中, 随深度局部放大的预测曲线中, 红色曲线也更加贴合黑色曲线的总体走势。图 15 为 Gu93 的测井解释图, 可以更好地观测出井下的各条测井曲线随深度变化的走势与关系。



(a) 两条复原曲线与其真实值之间的相关性



(b) 曲线 AC 复原结果的定量误差



(c) 曲线 DEN 复原结果的定量误差

图 14 重构定量结果可视化

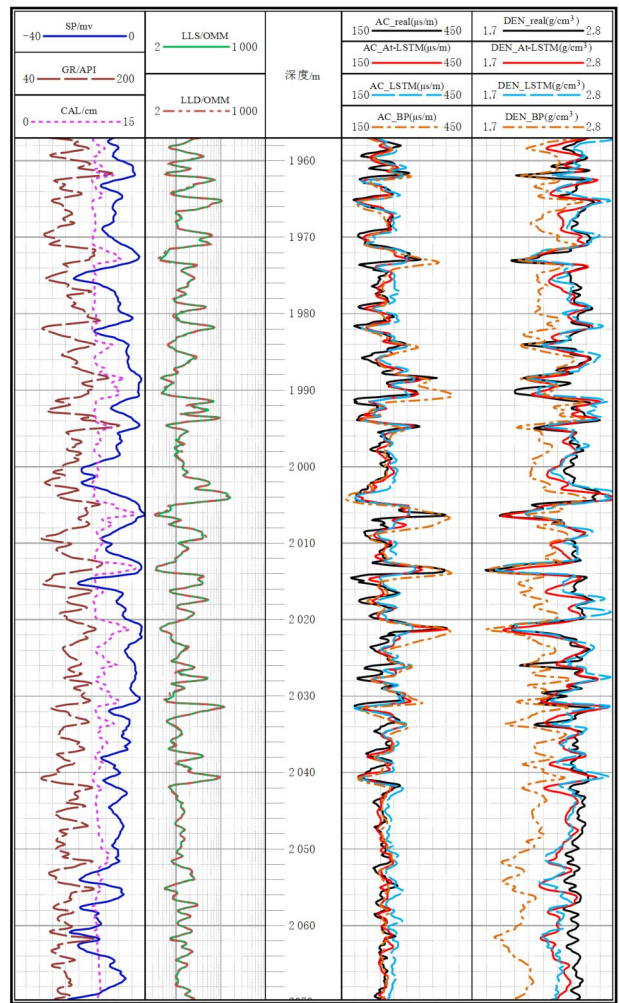


图 15 Gu93 井的测井解释图

### 5 结语

本文提出了一种 LSTM-Attention 预测模型, 并结合 EMD-VMD 分解后的模态分量特征来重构缺失的测井曲线。对缺失的 AC 和 DEN 曲线的预测结果表明, LSTM-Attention 模型重构出的曲线与真实曲线之间的 PCCs 分别可以达到 86.8% 和 74.8%, 高于 LSTM 与 BP 传统模型, 且具有更小的误差, 所以本文所提模型可以更好地重构缺失段曲线。在 LSTM-Attention 预测模型的基础上加入去除冗余信号的模态分量特征后, 两条测井曲线的预测精度进一步提升, AC 和 DEN 曲线的 PCCs 分别提升到了 88.2% 和 78.8%, 误差也降低了。综上所述, 本文所提的 LSTM-Attention 综合模型相比传统神经网络模型可以保持较高的预测精度和效率, 为油气勘探的后续进程提供了有力的数据保证。

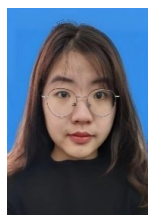
### 参考文献

[1] 张美玲, 丛琳, 张士奇, 等. 大庆外围扶杨油层组低孔渗

- 储层饱和度计算方法[J]. 大庆石油学院学报, 2011, 35(6): 9-12, 27, 122.
- ZHANG M L, CONG L, ZHANG S Q, et al. Calculation of saturation in Daqing Fuyang low permeability and porosity reservoirs[J]. Journal of Daqing Petroleum Institute, 2011, 35(6): 9-12, 27, 122. (in Chinese)
- [2] 王俊瑞, 梁力文, 邓强, 等. 基于多元回归模型重构测井曲线的方法研究及应用[J]. 岩性油气藏, 2016, 28(3): 113-120.
- WANG J R, LIANG L W, DENG Q, et al. Research and application of log reconstruction based on multiple regression model[J]. Lithologic Reservoirs, 2016, 28(3): 113-120. (in Chinese)
- [3] ZHANG Z W, YI W, KONG L J. Target detection and localization using multi-frame information for noncoherent MIMO radar[C]//IET International Radar Conference. London: IET, 2016: 1-6.
- [4] BANCHS R, JIMÉNEZ J R, DEL PINO E. Nonlinear estimation of missing logs from existing well log data[C]//SEG Technical Program Expanded Abstracts 2001. Houston: Society of Exploration Geophysicists, 2001: 598-600.
- [5] 吕晶, 谢润成, 周文, 等. LM-BP 神经网络在泥页岩地层横波波速拟合中的应用[J]. 中国石油大学学报(自然科学版), 2017, 41(3): 75-83.
- LÜ J, XIE R C, ZHOU W, et al. Application of LM-BP neural network in simulation of shear wave velocity of shale formation[J]. Journal of China University of Petroleum (Edition of Natural Science), 2017, 41(3): 75-83. (in Chinese)
- [6] SHI Y L, WANG Y T, LIU X J, et al. Prediction of oil and gas distribution in K area by inheritance classification[J]. IOP Conference Series: Earth and Environmental Science, 2021, 660(1): 012135.
- [7] 侯斌, 桂志先, 许辉群, 等. 应用多属性神经网络方法预测油气[J]. 岩性油气藏, 2010, 22(3): 118-120.
- HOU B, GUI Z X, XU H Q, et al. Application of multi-attribute and neural network method to hydrocarbon reservoir prediction[J]. Lithologic Reservoirs, 2010, 22(3): 118-120. (in Chinese)
- [8] 李守巨, 刘迎曦. 基于分形理论的煤层瓦斯渗透率参数估计方法[J]. 西安石油大学学报(自然科学版), 2007(2): 113-115, 180.
- LI S J, LIU Y X. Estimation method of the methane gas permeability in coal seam based on fractal theory[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2007(2): 113-115, 180. (in Chinese)
- [9] ARAKAWA K, KROTKOV E. Fractal surface reconstruction with uncertainty estimation: Modeling natural terrain [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 1993: 314-320.
- [10] REZAEI M, FAZEL ZARANDI M H. Facility location via fuzzy modeling and simulation[J]. Applied Soft Computing, 2011, 11(8): 5330-5340.
- [11] ALMARAASHI M, JOHN R. Tuning of Type-2 fuzzy systems by simulated annealing to predict time series[C]//Proceedings of the World Congress on Engineering 2011 Vol II. London: WCE, 2011: 1-5.
- [12] HA H, OLSON J R, BIAN L, et al. Analysis of heavy metal sources in soil using kriging interpolation on principal components[J]. Environmental Science & Technology, 2014, 48(9): 4999-5007.
- [13] BATEMAN R M. Openhole Log Analysis and Formation Evaluation[M]. Richardson: Society of Petroleum Engineers, 2012.
- [14] ASQUITH G B, KRYGOWSKI D, HENDERSON S, et al. Basic Well Log Analysis[M]. Tulsa: American Association of Petroleum Geologists, 2004.
- [15] ROLON L, MOHAGHEGH S D, AMERI S, et al. Using artificial neural networks to generate synthetic well logs [J]. Journal of Natural Gas Science and Engineering, 2009, 1(4/5): 118-133.
- [16] 杨志力, 周路, 彭文利, 等. BP 神经网络技术在声波测井曲线重构中的运用[J]. 西南石油大学学报(自然科学版), 2008, 30(1): 63-66, 13.
- YANG Z L, ZHOU L, PENG W L, et al. Application of bp neural network technology in sonic log data rebuilding[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2008, 30(1): 63-66, 13. (in Chinese)
- [17] 张东晓, 陈云天, 孟晋. 基于循环神经网络的测井曲线生成方法[J]. 石油勘探与开发, 2018, 45(4): 598-607.
- ZHANG D X, CHEN Y T, MENG J. Synthetic well logs generation via recurrent neural networks[J]. Petroleum Exploration and Development, 2018, 45(4): 598-607. (in Chinese)
- [18] 王俊, 曹俊兴, 尤加春. 基于 GRU 神经网络的测井曲线重构[J]. 石油地球物理勘探, 2020, 55(3): 510-520, 468.
- WANG J, CAO J X, YOU J C. Reconstruction of logging traces based on GRU neural network[J]. Oil Geophysical Prospecting, 2020, 55(3): 510-520, 468. (in Chinese)
- [19] 郑艳秋. 基于经验模式分解和神经网络的短期电力负荷预测方法研究[D]. 重庆: 重庆大学, 2009.
- ZHENG Y Q. Research on Short-term Power Load Forecasting Based on Empirical Mode Decomposition and Artificial Neural Network[D]. Chongqing: Chongqing University, 2009. (in Chinese)
- [20] 刘建华, 李锦程, 杨龙月, 等. 基于 EMD-SLSTM 的家庭短期负荷预测[J]. 电力系统保护与控制, 2019, 47(6): 40-47.
- LIU J H, LI J C, YANG L Y, et al. Short-term household load forecasting based on EMD-SLSTM[J]. Power System Protection and Control, 2019, 47(6): 40-47. (in Chinese)
- [21] 刘兴杰, 米增强, 杨奇逊, 等. 一种基于 EMD 的短期风速

- 多步预测方法[J]. 电工技术学报, 2010, 25(4): 165-170.
- LIU X J, MI Z Q, YANG Q X, et al. A novel multi-step prediction for wind speed based on EMD[J]. Transactions of China Electrotechnical Society, 2010, 25(4): 165-170. (in Chinese)
- [22] AN X L, YANG J J. Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy[J]. Transactions of the Institute of Measurement and Control, 2016, 38(3): 282-292.
- [23] ALI M, KHAN A, REHMAN N U. Hybrid multiscale wind speed forecasting based on variational mode decomposition[J]. International Transactions on Electrical Energy Systems, 2018, 28(1): e2466.
- [24] 陈强伟, 蔡文皓, 牛春光, 等. 基于VMD的APF谐波检测算法[J]. 电力科学与技术学报, 2018, 33(1): 120-124. CHEN Q W, CAI W H, NIU C G, et al. A APF harmonics detection method based on VMD[J]. Journal of Electric Power Science and Technology, 2018, 33(1): 120-124. (in Chinese)
- [25] 王俊, 李霞, 周昔东, 等. 基于VMD和LSTM的超短期风速预测[J]. 电力系统保护与控制, 2020, 48(11): 45-52. WANG J, LI X, ZHOU X D, et al. Ultra-short-term wind speed prediction based on VMD-LSTM[J]. Power System Protection and Control, 2020, 48(11): 45-52. (in Chinese)
- [26] 李鑫, 王娟, 邱亚, 等. 基于VMD的混合储能容量优化配置[J]. 太阳能学报, 2022, 43(2): 88-96. LI X, WANG J, QIU Y, et al. Optimal allocation of hybrid energy storage capacity based on variational mode decomposition[J]. Acta Energetica Sinica, 2022, 43(2): 88-96. (in Chinese)
- [27] 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 Well Log Completion and Generation Based on Machine Learning[D]. Beijing: Peking University, 2020. (in Chinese)
- [28] 王薪凯, 于忠清. 基于LSTM神经网络模型的泵站能耗预测[J]. 青岛大学学报(自然科学版), 2022, 35(1): 11-18. WANG X K, YU Z Q. Energy consumption prediction of pumping station based on LSTM neural network model [J]. Journal of Qingdao University (Natural Science Edition), 2022, 35(1): 11-18. (in Chinese)
- [29] SUBAKAN C, RAVANELLI M, CORNELL S, et al. Attention is all You need in speech separation[C]//ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2021: 21-25.
- [30] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6450-6458.
- [31] 徐延勇, 邹冠贵, 曹文彦, 等. 测井曲线标准化方法对比研究及应用[J]. 中国煤炭地质, 2013, 25(1): 53-57. XU Y Y, ZOU G G, CAO W Y, et al. Comparative study and application of logging standardization methods[J]. Coal Geology of China, 2013, 25(1): 53-57. (in Chinese)
- [32] 卢元磊, 何佳洲, 安瑾. 目标预测中的野值剔除方法研究[J]. 计算机与数字工程, 2013, 41(5): 722-725. LU Y L, HE J Z, AN J. Outlier-discrimination algorithm in target-prediction[J]. Computer & Digital Engineering, 2013, 41(5): 722-725. (in Chinese)
- [33] 曹志民, 路成辉, 刘爽, 等. 基于二分空间拉依达法的野值点剔除研究[J]. 化工自动化及仪表, 2018, 45(2): 137-140, 171. CAO Z M, LU C H, LIU S, et al. Research on outliers elimination based on binary-split space pauta method[J]. Control and Instruments in Chemical Industry, 2018, 45(2): 137-140, 171. (in Chinese)
- [34] 刘明阳. 基于LSTM模型的现代有轨电车到站时间预测方法研究[D]. 南京: 东南大学, 2020. LIU M Y. Research on the Arrival Time Prediction Method of Modern Streetcar Based on the LSTM Model[D]. Nanjing: Southeast University, 2020. (in Chinese)

#### 作者简介



刘梦女, 1997年生, 江苏徐州人. 东北石油大学硕士研究生. 主要研究方向为机器学习与多源异构时序信号处理等.

E-mail: m18362087659@163.com



韩建男, 1976年生, 黑龙江大庆人. 东北石油大学物理与电子工程学院教授, 博士生导师. 主要研究方向为油田智能感知、油田大数据和智能检测等.

E-mail: hanjian@nepu.edu.cn



曹志民男, 1980年生, 北京人. 东北石油大学物理与电子工程学院副教授, 硕士生导师. 主要研究方向为多源异构大数据模式识别与人工智能处理、勘探开发大数据分析及应用、计算机视觉等.

E-mail: caozhimin@nepu.edu.cn



刘兴斌男, 1964年生, 黑龙江绥化人. 东北石油大学物理与电子工程学院教授, 博士生导师. 主要研究方向为水平井油水两相流测量、采油井油气水三相流测量传感器研究等.

E-mail: liuxingbin256321@163.com