

基于自编码器的域适应时空测井曲线预测模型

刘成丽, 朱晨光

(长江大学地球物理与石油资源学院, 武汉 430100)

摘要: 以往研究中, 测井曲线特征提取的不完整和模型构建较为简单, 导致孔隙度预测精度受限。为提升预测精度, 结合自编码器、长短期记忆网络和注意力机制, 构建 AE-LSTM-AT(自编码器-长短期记忆网络-注意力机制)模型。AE(自编码器)将源域数据和目标域数据的特征分布统一到同一空间, 以降低因数据分布差异而引起的量纲变化对模型的干扰; 改良后的 LSTM(长短期记忆网络)在降低了参数数量的同时, 也增强了远距离时间步的特征影响, 减少信息污染; 而 Attention(注意力)机制的引入动态计算每个时间步的注意力权重, 从而更精准地聚焦关键特征, 提高模型在处理序列数据时的性能和表现。设立对照组 MLP(多层感知器)和 LSTM, 进行 4 组对比实验, 实验证明本文的模型结构在长期预测及跨域预测问题上具有较优效果。

关键词: 自编码器; 测井曲线; 长短期记忆网络; 注意力机制; 孔隙度预测

中图分类号: P631 **文献标志码:** A **文章编号:** 1671-1807(2025)03-0013-07

在石油工程勘探领域中, 测井数据极具价值, 因为它们提供了对石油储层存储和流动能力的描述, 以便于提高对地下储层的认识^[1]。孔隙度是描述地下储层中岩石孔隙空间的重要指标, 直接影响着地下流体的储存和运移能力^[2]。孔隙度的精确获取对油气藏开发方案的有效制定提供重要参考依据^[3]。测井数据是目前所能获得的分辨率最高、连续性最好的地质数据, 利用测井数据进行孔隙度等储层参数的预测一直是学者研究的重要课题^[4]。

机器学习(machine learning)的发展异常迅速, 已经逐步在很多领域有着令人瞩目的成就, 因其出色的非线性拟合能力, 在油气勘探领域迅速展开。这些方法被广泛用于预测储层参数。然而, BP(back propagation, 反向传播)神经网络存在一些固有问题, 包括收敛速度慢、易陷入局部最优解以及网络层结构的限制^[5]。深度学习逐渐成为机器学习领域的研究热点, 在许多工程领域之中应用广泛且有很多显著的成果进展。杨柳青等^[6]利用深层卷积神经网络预测模型, 将深度学习技术引入储层参数预测领域。使用密度、声波速度、自然伽马和泥质含量这 4 种测井参数对卷积神经网络进行训练, 得到的孔隙度预测结果与实际值高度吻合。刘俊等^[7]利用双向长短期记忆(BiLSTM)神经网络充分利用其处理序列化结构数据时前后两个方向的优势。

通过多次训练迭代优化模型, 建立敏感参数[CNL(补偿中子)、DEN(密度)、DT(声波)、GR(自然伽马)]与孔隙度之间的非线性映射关系。最终, 得到的 BiLSTM 模型在孔隙度预测方面表现优于长短期记忆(long short term memory, LSTM)、循环神经网络(recurrent neural network, RNN)和深度神经网络(deep neural networks, DNN)模型。安鹏等^[8]采用 LSTM 循环神经网络进行孔隙度预测, 并与全连接深度神经网络进行了比较。研究结果显示, LSTM 循环神经网络在预测精度和稳定性方面明显优于全连接深度神经网络, 展现出其在孔隙度预测任务之中的显著优势。

对前人的研究成果进行总结分析, 本文创新性提出 AE-LSTM-AT(auto-encoder-long short-term memory network-attention mechanism, 自编码器-长短期记忆网络-注意力机制)模型, 在模型结构的首部引入 AE 结构, 通过降低源域数据和目标域数据的分布差异, 以提高模型参数权重在训练集和测试集上的一致性。自编码器(auto-encoder, AE)的引入不仅可以得到高维度潜在特征表示, 而且可以降低量纲变化对模型的干扰。改良 LSTM 的优点在于处理测井数据这类数据时, 较传统 LSTM 能保留更多的有效特征。Attention 机制为每个时间步产生一个注意力权重向量, 用于加权输入序列的特征表示, 模型能

收稿日期: 2024-08-28

作者简介: 刘成丽(1998—), 女, 黑龙江佳木斯人, 硕士研究生, 研究方向为深度学习与测井资料综合解释; 朱晨光(1999—), 男, 湖北荆州人, 硕士研究生, 研究方向为测井解释与人工智能。

够有选择地关注与当前任务相关的信息,为改良版 LSTM 输出特征中的当前深度特征分量聚焦更多的深度方向特征信息。AE-LSTM-AT 模型使得生成的数据具有连贯性与更高的准确度。

1 方法原理

1.1 自编码器(AE)

自编码器是一种无监督学习的神经网络模型^[9],其主要任务是捕获输入数据中的潜在特征,从而能够有效地重构原始数据。其设计旨在通过学习数据的内部表示来提高数据表达和处理的效率。其优势在于可以同时学习原始数据中的线性和非线性信息,通过使用神经网络结构和激活函数进行非线性映射,更好地捕捉数据的复杂特征^[10]。自编码器主要由编码器(Encoder)以及解码器(Decoder)这两部分组成,编码器用于将输入数据压缩到潜在特征空间^[11],解码器则是将潜在的空间数据还原为最接近输入数据的状态^[12]。

多层感知器(multilayer perceptron, MLP)作为编码器的基本组件,通过多层神经网络的堆叠,可以实现复杂地质特征的有效提取。本文采用多层感知器作为解码器的基本组件,通过反向传播算法优化解码器的参数,实现对潜在特征的准确还原。AE-LSTM-AT 模型首先利用自编码器对密度、声波、中子和自然伽马 4 种测井数据进行特征学习,通过训练自编码器,可以得到测井输入参数的高维度潜在特征表示。以此将源域数据和目标域数据的特征分布统一到同一空间,以降低因数据分布差异而引起的量纲变化对模型的干扰。

1.2 长短期记忆神经网络(LSTM)

LSTM 是 RNN 的一个变体网络。LSTM 具有和 RNN 一样的链式结构,能提取数据的上文信息与数据本身的相关性,用于具有时序结构数据的处理。针对 RNN 结构的梯度爆炸和长期依赖的问题,LSTM 通过引入记忆单元的门结构进行了有效的解决。使用存储单元和门来控制通过网络的信息流,使它们能够在很长一段时期内保持信息,并避免因梯度问题而丢失信息,这使得它们非常适合于分析顺序数据和基于长期相关性进行预测^[13]。

LSTM 有 3 个门结构,分别为遗忘门、输入门和输出门,其中遗忘门控制上一个记忆单元输出信息哪些被遗弃,输入门控制哪些新的信息被记忆在当前记忆单元,输出门控制当前记忆单元哪些信息被输出,设组件二的历史特征注意力交互层输出结

果为 x_i , 其集合为 X , 即

$$X = \{x_1, x_2, \dots, x_{d-1}\} \quad (1)$$

遗忘门的计算表示为

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), x_t \in X \quad (2)$$

输入门的计算表示为

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), x_t \in X \quad (3)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), x_t \in X \quad (4)$$

输出门的计算表示为

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), x_t \in X \quad (4)$$

式中: σ 为激活函数; W, b 为权重和偏置矩阵; h_{t-1} 为前一时刻的隐藏状态; x_t 为当前时刻输入值; f_t 为遗忘门输出; i_t 为输入门输出; o_t 为输出门输出; \tilde{C}_t 为临时记忆状态,即输入门控制的新信息。

每个记忆单元的输出为 h_t 和 C_t , 而 h_t 又由当前记忆状态 C_t 决定,用公式表示为

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (6)$$

$$h_t = o_t \tanh C_t, t \in \{1, 2, \dots, d-1\} \quad (7)$$

LSTM 可以提取测井参数在深度方向的序列特征,根据图 1 的热力图发现时序上存在长期特征的回旋现象,前期特征对末端的影响比中间大,所以在第 2 层 LSTM 中删除部分不重要的序列特征,以保留前中期序列的特征。AE-LSTM-AT 模型中改良版 LSTM 通过测井参数之间的先验规律,进行一次 LSTM 网络后,对输出值按规律筛选后进行拼接,降低参数量的同时,增强了远距离时间步的特征影响,减少信息污染。

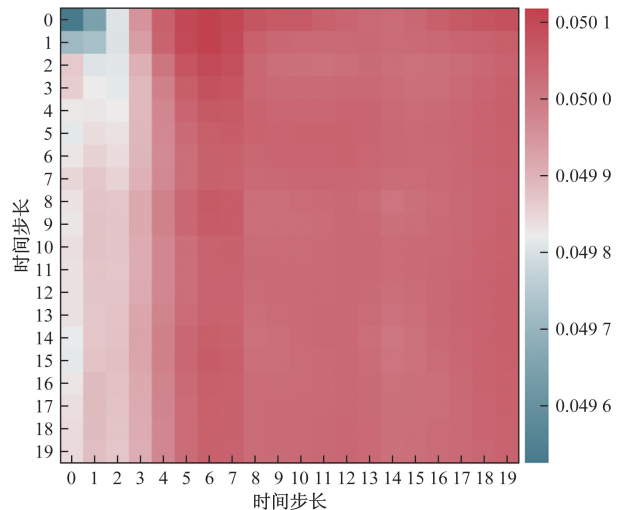


图 1 热力图

1.3 注意力机制(attention mechanism)

在处理序列数据时,注意力机制是一项关键技术,它使模型有选择性地聚焦于特定部分,以提高处理效率和准确性。通过这种机制,使模型更有针

对性,以应对数据的复杂性和多样性。在深度学习中,尤其是在处理序列数据时,注意力机制的引入能够使模型更加灵活地对输入数据进行处理。注意力机制的原理是通过学习或计算得到每个时间步的注意力权重,这些权重可以通过不同的方式计算。计算公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (8)$$

式中: \mathbf{Q} 为查询矩阵; \mathbf{K} 为键矩阵; \mathbf{V} 是值矩阵; d_k 为键的维度。

LSTM层仅以时间步长为间隔来捕捉序列信息的相关性,不利于深度序列的大跨度信息记忆^[14]。AE-LSTM-AT模型中,引入注意力机制位于改良版LSTM层之前,为每个时间步产生一个注意力权重向量,用于加权输入序列的特征表示。它根据输入数据和目标数据之间的相似度来分配权重,这个权重向量决定了每个时间步的输入对当前时间步的重要性,权重越大,表示输入数据对预测目标越重要^[13]。使模型能有选择地关注与当前任务相关的信息,为改良版LSTM输出特征中当前深度特征分量聚焦更多的深度方向特征信息。

1.4 本文模型

AE-LSTM-AT模型将AE、改良的LSTM以及Attention机制进行优势融合,能更有效地捕捉数据之间的复杂关系,实现孔隙度与测井数据之间映射关系的精确建模和精准预测。本文模型根据模块的作用,可划分为3个组件,如图2所示。

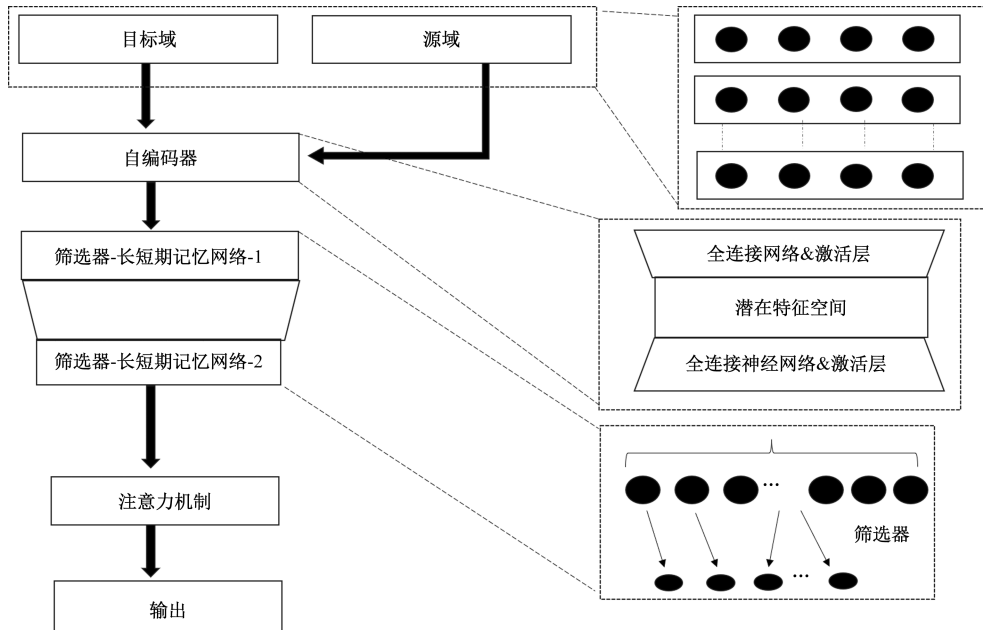


图2 模型结构

2 实例应用

2.1 实验数据集与数据预处理

实验数据集由某区域的A井和B井中真实测井数据组成,包含1578组测井数据样本。为了进行域自适应的孔隙度预测实验,将A井和B井的数据分别划分为源域数据和目标域数据。从A井中选取深度范围4165~4279.7m的数据作为源域数据集,共包含1147组测井数据样本。从B井中选取深度范围4166.9~4210m的数据作为目标域数据集,共包含431组测井数据样本。

在进行实验之前,先从原始数据中提取出与孔隙度相关的特征,选取密度(DEN)、声波时差(AC)、补偿中子(CNL)和自然伽马(GR)4种测井变量特征作为模型的输入。AC、DEN、CNL和GR测井曲线不仅反映储层的物性,亦映射了储层的含气性^[15]。对原始数据进行数据预处理,通常会丢弃测井数据中的缺失或异常值,以确保数据的质量和准确性。由于数据之间存在量纲上的差异,会导致不同的数据有较大差距,进而影响模型的预测性能。为了解决这一问题,进而对每个特征进行归一化处理,以确保它们都调整到相同的数值范围内,公式为

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

式中: \bar{x} 为归一化后的测井数据, $\bar{x} \in [0, 1]$; x 为原始数据; x_{\min} 和 x_{\max} 为测井样本数据的最小值和最大值。

2.2 准确度评估方法

为了评估 AE-LSTM-AT 模型在准确性和先进性方面的表现,采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)和决定系数(R^2)作为评估模型预测能力的指标。MAE 具有较强的鲁棒性,对异常值的影响较小,更能全面反映模型的整体误差。RMSE 对异常值更为敏感,更突出了这些值对模型预测准确性的影响。MAE 和 RMSE 数值越小,模型效果越好。 R^2 取值范围介于 $0\sim 1$, R^2 越接近 1,则说明模型能够更有效地解释孔隙度数据的变化,表明模型的拟合效果越佳。将这 3 种指标结合可以更精准地评估模型的预测性能,计算公式分别为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (12)$$

式中: y_i 、 \hat{y}_i 分别为实际孔隙度值和预测孔隙度值; n 为样本数。

2.3 实验分析

2.3.1 域适应实验

为了验证自编码器在孔隙度预测中的效果,同时避免其只在本文模型下效果明显,选取同一多层感知机神经网络(MLP)模型作为对比,以评估添加自编码器后的性能提升情况。设计了两个实验条件:①基准模型(Baseline):仅使用多层感知机神经网络(MLP)模型进行孔隙度预测;②添加自编码器的模型(AE-MLP):在基准模型的基础上,添加自编码器来学习数据的高级特征表示。对每个实验条件下的模型输入 A 井数据集进行训练,使用 B 井数

据对模型预测效果进行对比,得出孔隙度预测精度对比(表 1)。通过表 1 可以看出,添加自编码器的模型 MAE 和 RMSE 分别低于未添加的基准模型 0.314 4 和 0.294 8,误差更小,并且 R^2 高 0.278 7,可见自编码器在孔隙度预测中体现出更好的性能。根据不同模型的预测数值绘制相应的孔隙度预测曲线,如图 3 所示。图 3 展示了孔隙度曲线的拟合效果,添加了自编码器的模型预测曲线与孔隙度实际值曲线相关性更高。基准模型参数的最终权重取决于训练井的数据分布,而实际情况中地质问题上具有较强非均质,同井不同段,不同井之间的数据分布存在较大差异,基准模型无法学习到目标域的数据分布结构,故存在较大的误差,而本文提出利用先验知识,通过自编码器将源域数据和目标域数据映射到同一特征域中,降低了跨域问题上数据分布的差异,故而能较明显地提高模型的准确度。

表 1 孔隙度预测精度对比

模型	MAE/%	RMSE/%	R^2
AE-MLP	0.890 3	1.116 2	0.590 9
MLP	1.204 7	1.411 0	0.312 2

2.3.2 LSTM 改良实验

为对比本文提出的改良版 LSTM 和传统 LSTM 在孔隙度预测上的性能表现,选用 A 井数据集进行数据训练,B 井数据集进行实验预测。改良版 LSTM 和传统 LSTM 的超参数配置保持一致,以消除其他因素的干扰。分别使用两种模型对测井数据进行训练和测试,得到预测精度对比结果(表 2)。表 2 中改良版 LSTM 的 MAE 和 RMSE 值略小于传统 LSTM, R^2 略大于传统 LSTM,表明两种网络模型在预测上虽表现接近,但改良版 LSTM 的误差更小,准确性更高。根据预测数值绘制出预测对比图,如图 4 所示,通过图 4 可以看到改良版

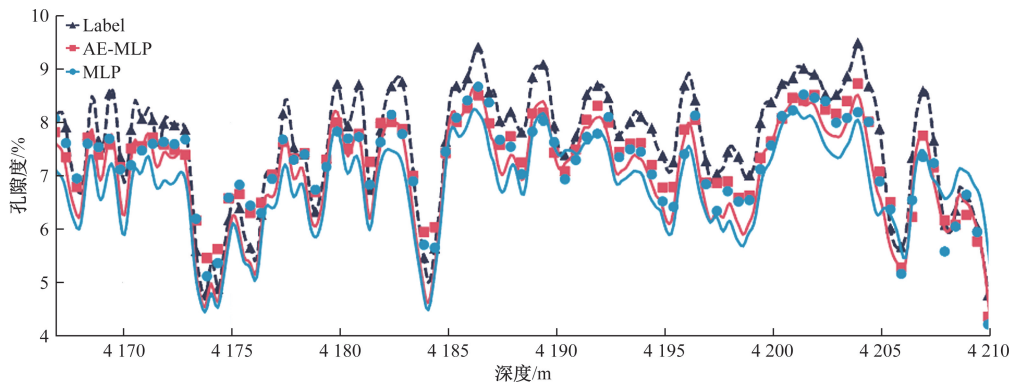


图 3 不同模型孔隙度预测曲线

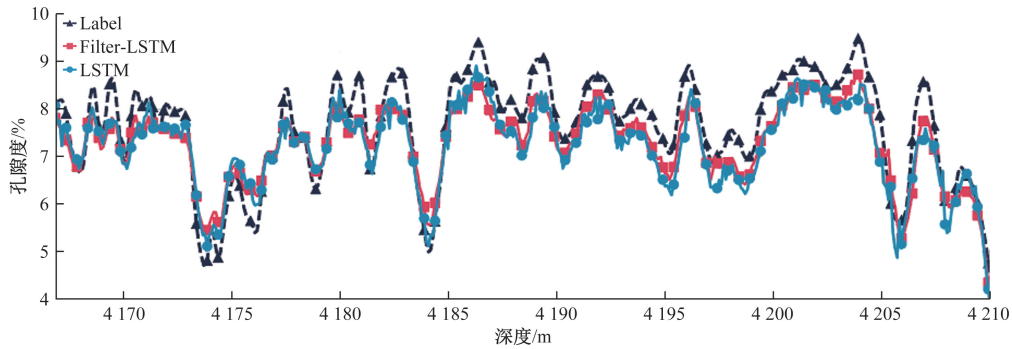


图 4 不同模型孔隙度预测对比

表 2 孔隙度预测精度对比

模型	MAE/%	RMSE/%	R^2
Filter-LSTM	0.706 9	0.892 2	0.690 5
LSTM	0.754 6	0.967 0	0.626 4

LSTM 的总体稳定性更强,传统 LSTM 曲线波动较大,虽二者较为接近,但是改良版 LSTM 拟合精度高于传统 LSTM,在预测精度上有着明显的优势。测井参数较单一,数值具有较高信息量,模型在增强特征提取能力的同时,会因为更加复杂的模型参数从而导致污染测井参数的信息,AE-LSTM-AT 模型根据测井参数的深度规律,改良了 LSTM 模块,保留测井数据的更多有效特征,增加模型特征提取能力的同时,降低模型复杂度,避免信息污染。

2.3.3 同井对比实验

进行一组消融实验来分析 AE-LSTM-AT 模型中各个模块对最终模型性能的贡献度。通过逐步移除模型中的某些模块,评估每个模块对 AE-LSTM-AT 模型性能的影响。设计了以下几个消融实验条件:完整模型、移除自编码器模块、移除 LSTM 模块、移除注意力机制模块。

对每个实验条件下的模型就 A 井数据集进行训练,并使用 B 井数据集进行预测以此得到各模块的最终贡献对比,对比结果见表 3。从表 3 中可以看到移除模块中各个部分都会对最终模型的预测性能产生影响,导致最终的预测性能降低。移除自编码器模块后,MAE 和 RMSE 分别比完整模型有所增加,误差变高,精度降低。因为自编码器使模型可以实现域适应并且克服了非均质性所以提升了模型的预测效果;移除改良版 LSTM 模块后,MAE 和 RMSE 分别比完整模型增加了 0.809 5 和 0.888 7,有着较大差别,也证明了改良版 LSTM 使模型更好地捕捉序列数据中的时间依赖关系,保留数据更多有效特征;移除注意力机制模块后,MAE

表 3 各模块贡献度对比

组成部分	部分选择			
Auto Encoder	✓		✓	✓
Filter-LSTM	✓	✓		✓
Self-Attention	✓	✓	✓	
MAE/%	0.355 7	0.538 8	1.165 2	0.621 5
RMSE/%	0.578 3	0.804 1	1.467 0	0.924 0
R^2	0.866 1	0.724 6	0.350 4	0.702 8

注:Self-Attention 本身的特征提取能力较弱,消融部分仅用线性层进行连接。

和 RMSE 也有着一定的涨幅,模型无法很好地关注输入序列中的重要特征所以导致预测误差变大。消融实验结果表明,AE-LSTM-AT 模型中的每个模块都对最终模型性能发挥有着重要的贡献,它们的组合使模型融合 3 种技术方法的优势于一体,能够在孔隙度预测中取得最好的效果。

2.3.4 跨井对比实验

为了进一步比较 AE-LSTM-AT 模型与其他模型在预测孔隙度方面的优越性,使用相同数据量对多种模型进行训练,并进行了孔隙度预测精度的对比实验。根据表 4 不同模型在 B 井上的量化指标的结果显示,AE-LSTM-AT 模型的 MAE 为 0.355 7, RMSE 为 0.578 3,在预测偏差和相关性方面明显优于 MLP 和 LSTM 模型。 R^2 为 0.866 1,证明 AE-LSTM-AT 模型可以更准确、更可靠地预测储层参数。MLP 的 MAE 和 RMSE 值最高,从图 5 可以看出,在深度 4 185~4 205 m 段误差最大,因为 AE-LSTM-AT 模型和 LSTM 均引入时间特征,可以有效提升精度。AE-LSTM-AT 模型的 MAE 和 RMSE 指标比 LSTM 分别低 0.398 9 和 0.388 7,从图 5 也可以看出,AE-LSTM-AT 模型能较好地拟合原始孔隙度曲线,说明 AE-LSTM-AT 模型相较于原始的 LSTM 模型保留更多数据的有效特征,具有更优秀的预测性能。

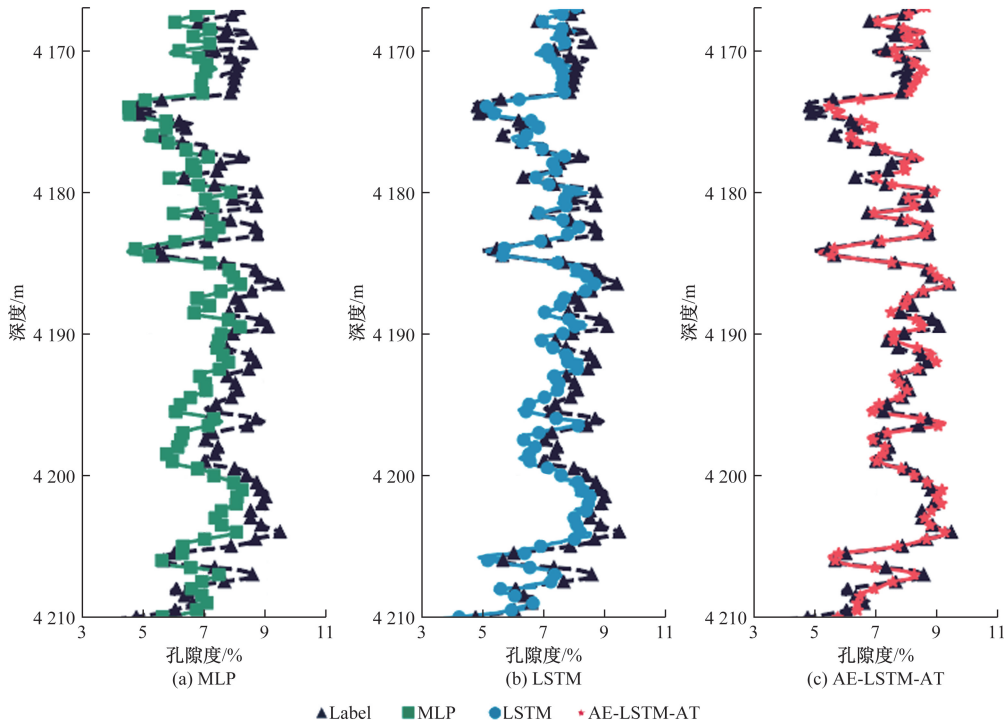


图 5 不同模型孔隙度预测结果

表 4 不同模型孔隙度预测精度对比

模型	MAE/%	RMSE/%	R^2
MLP	1.204 7	1.411 0	0.312 2
LSTM	0.754 6	0.967 0	0.626 4
AE-LSTM-AT	0.355 7	0.578 3	0.866 1

3 结论

本文通过大量实验,针对测井数据对传统 LSTM 结构进行了改良,并创新性地提出了一种域适应的时序模型结构 AE-LSTM-AT 模型。该模型的核心优势体现在以下几个方面。

(1)域适应的自编码器。AE-LSTM-AT 模型前端引入了自编码器结构,通过域适应技术,有效提取并调整测井数据的特征表示。这一模块显著提高了模型在不同测井环境中的泛化能力,使得模型能够在面对分布差异较大的数据时,依然保持较高的预测精度。

(2)改良的 LSTM 结构。针对测井数据的特点,AE-LSTM-AT 模型对传统 LSTM 进行了优化,使其更加适合处理复杂的地质序列信号。通过调整 LSTM 的内部结构和参数配置,模型在时间依赖性和特征提取能力上得到了增强,从而在测井数据处理上表现出更高的准确性。

(3)注意力结构。在 AE-LSTM-AT 模型的末端引入注意力机制,使模型能够更加有效地捕捉序列中的关键信息,并在全局上下文中权衡不同时间

步的贡献。该机制提升了模型的决策能力,使得其在应对长时间依赖关系和噪声较大的数据时,表现得更加稳健。

通过这些改进,本文提出的 AE-LSTM-AT 模型在测井数据预测任务中,表现出显著优于常见 MLP 和传统 LSTM 的性能,特别是在数据多样性和复杂性较高的情况下,展现出更强的适应性和鲁棒性。未来,随着测井数据处理需求的增加,本模型有望在更广泛的应用场景中发挥关键作用。

参考文献

- [1] ZENG L, REN W, SHAN L. Attention-based bidirectional gated recurrent unit neural networks for well logs prediction and lithology identification[J]. Neurocomputing, 2020, 414: 153-171.
- [2] 汪锐,李芳,刘仕友,等.岩石物理建模引导的低渗透层参数预测方法[J].煤田地质与勘探,2024,52(8):187-197.
- [3] 韩鸿来,刘之的,张志强,等.基于测井曲线包络面积预测油气藏产能[J].科学技术与工程,2024,24(15):6254-6262.
- [4] 段友祥,王言飞,孙歧峰.选择性集成学习模型在岩性-孔隙度预测中的应用[J].科学技术与工程,2020,20(3):1001-1008.
- [5] AN P, CAO D. The application of deep learning method in the prediction of shale content[C]//2018 International Geophysical Conference & Exposition Electronic Papers. Beijing: 中国石油学会(CPS),国际勘探地球物理学家学会(SEG),2018:4.

- [6] 杨柳青, 陈伟, 查蓓. 利用卷积神经网络对储层孔隙度的预测研究与应用[J]. 地球物理学进展, 2019, 34(4): 1548-1555.
- [7] 刘俊, 曹俊兴, 丁蔚楠, 等. 基于双向长短期记忆神经网络的储层孔隙度预测方法研究[J]. 地球物理学进展, 2022, 37(5): 1993-2000.
- [8] 安鹏, 曹丹平, 赵宝银, 等. 基于 LSTM 循环神经网络的储层物性参数预测方法研究[J]. 地球物理学进展, 2019, 34(5): 1849-1858.
- [9] 赵会芹. 基于不同空间分辨率遥感影像的滑坡提取方法研究[D]. 桂林: 桂林电子科技大学, 2023.
- [10] 高怡欣, 周云海, 燕良坤, 等. 基于稀疏自编码器降维和改进 BIRCH 聚类算法的台区拓扑关系辨识[J/OL]. 现代电子技术, 1-9[2024-07-02]. <http://kns.cnki.net/kcms/detail/61.1224.TN.20240319.1508.002.html>.
- [11] 麻健鹏. 基于弱磁信号的航空轴承早期故障诊断技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2022.
- [12] 魏晓飞, 李飞, 李素芳, 等. 一种基于自编码器的无监督漂移检测算法[J]. 河北冶金, 2024(4): 29-33.
- [13] GIAN A, RADHI M, AGUNG N, et al. Deep sequence model-based approach to well log data imputation and petrophysical analysis: a case study on the West Natuna Basin, Indonesia [J]. Journal of Applied Geophysics, 2023, 218: 105213.
- [14] YANG L Q, WANG S D, CHEN X H, et al. Deep-learning missing well-log prediction via long short-term memory network with attention-period mechanism[J]. Geophysics: Journal of the Society of Exploration Geophysicists, 2023, 88(1): 31-48.
- [15] 韩鸿来, 刘之的, 张志强, 等. 基于测井曲线包络面积预测油气藏产能[J]. 科学技术与工程, 2024, 24(15): 6254-6262.

Self-encoder-based Domain-adaptive Spatio-temporal Logging Curve Prediction Models

LIU Chengli, ZHU Chenguang

(School of Geophysics and Oil Resources, Yangtze University, Wuhan 430100, China)

Abstract: In previous studies, incomplete extraction of logging curve features and simpler model construction resulted in limited porosity prediction accuracy. In order to improve the prediction accuracy, the self-encoder, long and short-term memory network and the Attention mechanism were combined to construct the AE-LSTM-AT (auto-encoder-long short-term memory network-attention mechanism) model. the AE (self-encoder) unifies the feature distributions of the source domain data and the target domain data into the same space, in order to reduce the interference of the magnitude changes on the model due to the differences in data distribution, the modified LSTM(long short-term memory network) reduces the number of parameters while enhances the feature impact of distant time steps and reduces information pollution, and the introduction of the Attention mechanism dynamically calculates the attention weight of each time step, thus focusing on the key features more accurately and improving the performance and performance of the model in processing sequence data. a control group including MLP(multilayer perceptron machine) and LSTM was set up, and four sets of comparison experiments were conducted. It is proved that the model structure of has superior results in the problems of long-term prediction and cross-domain prediction.

Keywords: self-encoder; logging curve; long short-term memory network; attention mechanism; porosity prediction