

DOI: 10.16285/j.rsm.2024.0344

基于深度学习的岩石短长期本构模型及 灰砂岩变形预测

张 璠¹, 李书恒¹, 朱其志¹, 石玲玲¹, 邵建富^{1,2}

(1. 河海大学 土木与交通学院, 江苏 南京 210024; 2. 里尔大学, 法国 里尔)

摘 要: 岩石的短期和长期力学性能和变形特性对工程长期稳定与安全有着重要的影响。传统的本构模型难以统一描述不同岩石材料的短长期力学特性, 而基于深度学习方法的理论可在不引入其他弹塑性参数以及本构规律的情况下预测不同岩石的力学特性。长短期记忆 (long short-term memory, 简称 LSTM) 深度学习算法适用于处理具有时间序列的数据任务, 用于预测岩石短长期力学特性具有显著优势。通过引入 LSTM 算法, 分别根据三轴压缩加载路径和应力松弛随时间变化的规律构建序列数据, 建立了灰砂岩在常规三轴压缩以及应力松弛下的力学特性预测模型。与试验数据进行对比, 可以证明基于深度学习的岩石短长期力学预测本构模型的准确性。为进一步提升模型工程应用价值, 将 LSTM 本构模型嵌入到有限元法 (finite element method, 简称 FEM) 框架中进行数值实现, 并应用于灰砂岩变形特性的模拟。对比结果表明, LSTM-FEM 方法具有较好地预测岩石短长期变形特性的能力。

关 键 词: 短长期力学特性; 长短期记忆; 本构关系; 有限元法

中图分类号: TU 413

文献标识码: A

文章编号: 1000-7598 (2025) 01-0289-14

Short- and long-term rock constitutive model and gray sandstone deformation prediction based on deep learning method

ZHANG Jin¹, LI Shu-heng¹, ZHU Qi-zhi¹, SHI Ling-ling¹, SHAO Jian-fu^{1,2}

(1. College of Civil Engineering and Transportation, Hohai University, Nanjing, Jiangsu 210024, China; 2. University of Lille, Lille, France)

Abstract: The short and long-term mechanical properties and deformation characteristics of rocks significantly impact the engineering project's long-term stability and safety. Traditional constitutive models struggle to uniformly describe the short- and long-term mechanical properties of various rock materials. In contrast, deep learning methods can predict these properties without additional elastic or plastic parameters and constitutive laws. The long short-term memory (LSTM) deep learning algorithm is well-suited for time series data tasks and excels in predicting the short- and long-term mechanical properties of rocks. This study utilized the LSTM algorithm to construct sequence data based on the triaxial compression loading path and stress relaxation over time. A prediction model for the mechanical properties of gray sandstone under conventional triaxial compression and stress relaxation was established. Comparison with experimental data validates the accuracy of the deep learning-based short- and long-term rock mechanical prediction model. To enhance the model's engineering application value, the LSTM constitutive model was embedded into the finite element (FEM) framework for numerical implementation and applied to gray sandstone simulation. The comparison results show that the LSTM-FEM method better predicts the short- and long-term deformation characteristics of rocks.

Keywords: short- and long-term mechanical properties; long short term memory; constitutive relation; finite element method

1 引 言

岩石内部含有的微裂纹、节理和孔隙等缺陷被

称为岩石的初始损伤。在时间、外部荷载以及周围环境变化的作用下, 初始损伤会不断扩展、演化使得岩石结构发生重组, 改变岩石材料相关力学特性,

收稿日期: 2024-03-22

录用日期: 2024-07-08

基金项目: 中央高校基本科研业务费专项资金资助 (No.B230201059)。

This work was supported by the Fundamental Research Funds for the Central Universities (B230201059).

第一作者简介: 张璠, 男, 1990 年生, 博士, 副教授, 主要从事岩土材料本构理论和人工智能算法方面的研究, E-mail: chelseazhangjin@163.com

通讯作者: 朱其志, 男, 1979 年生, 博士, 教授, 主要从事岩石多尺度建模和数值模拟方法方面的研究, E-mail: qizhi_zhu@163.com

进而影响岩体工程的长期稳定性。因此,为保证工程的长期稳定性,不仅要考虑岩石的短期瞬时行为,还要考虑岩石的流变特性^[1-3]。国内外学者一般采用瞬时和时效本构模型来反映岩石的变形机制。

用于描述岩石流变特性的时效本构模型主要有经验模型、元件模型和基于损伤机制的流变模型^[4]共3类。经验模型在早期研究中被广泛应用,学者主要利用数学函数创建模型。张向东等^[5]使用幂次函数关系得出了泥岩的非线性蠕变方程。Yang等^[6]考虑围压与轴压作用,提出了一个指数方程来描述初期蠕变和稳定蠕变。经验模型的建立主要基于试验结果的总结,缺乏理论依据,且适用范围有限。元件模型具有物理意义明确,能直观地表现岩石蠕变特性等特点。Li等^[7]使用串联 Burgers 模型和塑性元件来描述层状复合岩石的蠕变特性。Zhou等^[8]利用非线性黏壶,提出微分形式的蠕变本构方程描述盐岩流变特性。徐卫亚等^[9]在五元件线性黏塑性流变模型中串联一新的非线性黏塑性体,建立了反映岩石加速流变特性的模型。这些组合元件模型往往公式复杂,参数常随变形、温度等条件变化。损伤模型的研究起始于 Kachanov^[10]研究蠕变破坏时首次使用“连续性因子”和“实际应力”等概念反映材料的损伤。针对岩石等准脆性材料,Dragon等^[11]建立起塑性损伤模型的理论框架。Shao等^[12]针对岩石材料建立了各向同性和各向异性的塑性损伤模型。Zhu等^[13-14]利用离散热力学公式开发了离散塑性损伤模型。

传统岩石本构模型的建立过程中,需要引入大量的假设来确定模型,且模型的数学形式复杂、参数标定困难、建立过程难度大,却常常不适用于不同类型和条件的岩石。

随着计算机技术的进步,利用深度学习算法确定材料力学行为的方法得到了发展。相比传统研究方式,深度学习无需任何力学假设。经过训练后的神经网络犹如“黑匣子”,用一组表征材料力学行为的完整数据对深层次神经网络进行训练,就能合理假设用深度学习方法获得的材料模型会逼近真实材料所遵循的全部力学定律。李守巨等^[15]采用误差反向传播(back propagation,简称BP)模型对大坝混凝土和基岩的弹性模量进行训练,预测结果与现场监测取芯结果基本一致。陈昌彦等^[16]采用BP网络建立了地质以及工程因素与安全系数和稳定状态之间的非线性学习模型。谭云亮等^[17]利用径向基(radial basic function,简称RBF)网络,基于岩石力学试验数据,确定相应的岩石的本构关系。然而,

传统人工神经网络由于不同时间步之间缺乏联系、误差累积、计算成本高等原因无法反映数据的历史相关性^[18]。为解决上述问题,计算机领域提出了循环神经网络(recurrent neural network,简称RNN),但RNN容易导致梯度消失,同时当处理数据长期依赖时精度也可能会大幅降低,输入输出序列不匹配,模型的数据共享所导致的缺失信息会对时序特征产生干扰。Hochreiter等^[19]提出长短时记忆网络模型(long short-term memory,简称LSTM)。同时,计算机科学领域也纷纷利用注意力机制改进LSTM模型^[20-21]。注意力机制的本质是实现信息处理资源的高效分配,当神经网络处理输入数据时,它可以自动学习哪些信息更加重要,然后在预测阶段对这些关键信息进行重点关注,以提高模型的准确性和效率。

综上所述,LSTM长短时记忆模型,是一种特殊的循环神经网络,其适用于处理具有时间序列的数据任务。与传统本构模型相比,基于LSTM建立的本构模型可以无需任何力学假设,不考虑模型参数,并适用于不同性质的岩石。此外,岩石的应力松弛特性是其长期力学性质中的一个重要分支。在应力松弛过程中,岩体内部应力大小和分布随时间变化。相较于其他深度学习模型,使用LSTM模型处理并预测这种岩石长期力学特性有着显著优势。注意力机制可以帮助解决长序列训练过程中的梯度消失问题,从而提高模型的训练效率和泛化能力。

因此,本文将探究如何利用LSTM模型和注意力机制,以预测灰砂岩的短期和长期力学特性,并将最优的LSTM深度学习本构模型在有限元程序中进行数值实现,应用于灰砂岩的模拟,以验证LSTM深度学习方法的有效性。

2 深度学习方法与数据准备

深度学习就是采用神经网络方法从有限的学习中学习具有一般性的规律,并利用这些规律对未知数据进行预测。本节将分别对所用神经网络方法和样本数据来源进行简要介绍。

2.1 深度学习的方法

2.1.1 长短时记忆网络

RNN是一种用于处理序列数据的神经网络。RNN能很好地反映材料历史相关特性。RNN的隐藏层的输入不仅包含当前时刻输入层的输出,还包含上一时刻隐藏层的输出。经研究发现,RNN在解决长序列问题时,训练过程中会出现长期记忆丧失。相较于RNN,LSTM添加了遗忘门、输入门、输出

门,如图 1 所示, A 代表一个 LSTM 单元, 每个单元通过“门控装置”很好地解决了长期依赖问题。其中: 遗忘门的作用是决定是否保留前一时间步 $t-1$ 的细胞状态 C_{t-1} , 即控制当前时刻 t 的细胞状态 C_t 中包含多少来自于之前的记忆内容。在 LSTM 中, 细胞状态 C_t 是一个内部存储单元, 包含了神经网络在特定时间步的长期记忆。输入门负责处理当前时刻的输入。LSTM 的关键组件是记忆细胞 (Cell State), 其功能是赋予 LSTM 能够在每个时间步自由选择记忆内容的的能力。从网络结构上看, LSTM 相较于传统 RNN, 有更复杂的输入。LSTM 的输入包括: 当前时刻 x_t 、前一时间步隐藏层的输出 h_{t-1} 以及前一时间步的记忆细胞状态 C_{t-1} 。其中, x_t 表示当前时刻 t 的输入向量, 代表输入数据的特征; h_{t-1} 表示前一时间步 $t-1$ 的隐藏状态, 包含到上一时间步为止的历史信息。

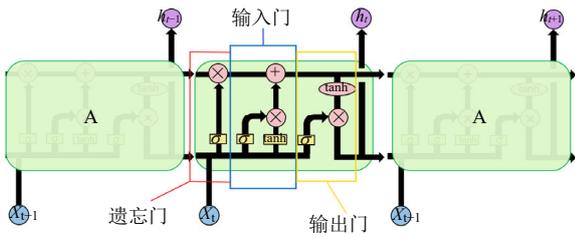


图 1 LSTM 网络按时间展开的结构

Fig.1 Structure of LSTM network unfolded in time

遗忘门的计算表达式如下:

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (1)$$

式中: f_t 为遗忘门的输出值, 取值范围 $[0,1]$, “0”代表完全遗忘前一时间步信息, “1”代表完全保留信息; W_f 、 U_f 为遗忘门的权重矩阵; b_f 为遗忘门的偏置项; $\sigma(*)$ 为 sigmoid 激活函数, 其数学表达式为 $\sigma(*) = 1/(1+e^{-*})$, 其中 e 是自然底数。

输入门的计算表达式如下:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (2)$$

式中: i_t 为当前输入 x_t 和 h_{t-1} 能保存到单元状态中的比例; W_i 、 U_i 为输入门的权重矩阵; b_i 为输入门的偏置项。

中间状态变量 C_t 更新计算表达式如下:

$$C_t = C_{t-1} \odot f_t + i_t \odot \tilde{C}_t \quad (3)$$

式中: \odot 为哈达玛积; \tilde{C}_t 为当前输入的单元候选状态, 且有

$$\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (4)$$

式中: W_c 、 U_c 为单元候选状态的权重矩阵; b_c 为单元候选状态的偏置项; $\tanh(*)$ 是激活函数, 定义为 $\tanh(*) = (e^* + e^{-*})/(e^* - e^{-*})$ 。

输出门的计算表达式如下:

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (5)$$

式中: o_t 为输出门的输出值; W_o 、 U_o 为输出门的权重矩阵; b_o 为输出门的偏置项; h_t 为当前时刻 t 的隐藏状态, 由前一时间步 $t-1$ 的隐藏状态 h_{t-1} 与当前时刻的输入 x_t 通过 sigmoid 激活函数得到的输出 o_t , 以及更新后的记忆细胞状态 C_t 经过 tanh 激活函数所得的结果共同作用产生:

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

从上述工作原理可以发现, LSTM 网络模型可以有效应用于解决历史相关性问题的, 且对模型训练过程即是对模型的参数 W 、 U 、 b 进行学习更新的过程。

目前 LSTM 已经被广泛应用于许多领域。在岩土工程中的应用有: Zhang 等^[22]使用 LSTM 预测了颗粒材料的循环剪切行为。Karapiperis 等^[23]将训练得到的神经网络本构模型运用于有限元模型中。Qu 等^[24]基于离散元数据探索了 3 类不同的颗粒材料本构训练策略。

2.1.2 注意力机制

LSTM 属于编码器-解码器 (encoder-decoder) 框架。该框架是深度学习领域中的一种研究模式。编码器作用是对输入序列进行编码, 通过非线性变换将输入序列 X_1, X_2, \dots, X_{t-1} 转化为语义向量, 解码器作用则是根据语义向量和之前生成的历史信息 Y_1, Y_2, \dots, Y_{t-1} 生成当前时刻 Y_t 。编码器和解码器之间唯一的联系就是语义向量 C 。上述生成过程用数学表达式记为

$$\left. \begin{aligned} Y_1 &= f(C) \\ Y_2 &= f(C, Y_1) \\ Y_3 &= f(C, Y_1, Y_2) \end{aligned} \right\} \quad (7)$$

由式 (7) 可见, 不论生成 Y_1 、 Y_2 、 Y_3 , 语义编码 C 均相同, 即生成 Y_t 过程中, 输入序列的任意隐藏状态对 Y_t 的影响力均相同。因此, 编码器-解码器框架为注意力不集中的分心模型。当输入序列长度较短时, 注意力对模型准确度的影响相对较小。当输入序列长度较长时, 就会导致解码时准确度下降。因此, 引入注意力模型一定程度上可以提高 LSTM 模型的适用范围。增加注意力模型的编码器-

解码器框架如图 2 所示, 其原理是自动搜索与 Y_i 相关的隐藏状态, 使其的影响力权重相应增大。

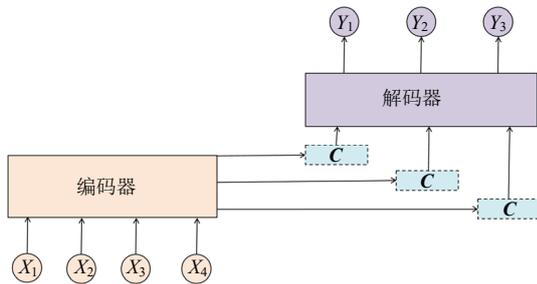


图 2 增加注意力机制的编码器-解码器模型
Fig.2 Encoder-Decoder model that adds attention mechanism

2.1.3 模型评价指标

对于回归模型, 性能的好坏主要体现在拟合曲线与真实曲线的误差。回归模型的性能评价指标能有效反映预测值与真实值之间的差异。单一评价指标对模型能力检验具有片面性。因此, 本文主要选取均方根误差 RMSE (root mean squared error)、平均绝对误差 MAE (mean absolute error)、决定系数 R^2 (coefficient of determination) 三个指标对模型预测结果的可靠性进行评价。

均方根误差 RMSE 是所有预测值与真实值的偏差的平方的平均值的平方根。取值范围是 $[0, +\infty)$, 当预测值与真实值完全吻合时, $RMSE=0$ 。RMSE 越大, 说明模型预测效果越差。RMSE 表达式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - y_i)^2} \quad (8)$$

式中: y 为模型预测值; y_i 为真实值; n 为样本容量。

平均绝对误差 MAE 是所有预测值与真实值的偏差的绝对值的平均值。通过取所有偏差的绝对值, 避免误差之间相互抵消。取值范围是 $[0, +\infty)$, 当预测值与真实值完全吻合时, $MAE=0$ 。MAE 越大, 说明模型预测效果越差。MAE 表达式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - y_i| \quad (9)$$

决定系数 R^2 是回归平方和占总离差平方和的权重。取值范围一般为 $[0, 1]$, R^2 越接近 1, 则说明模型预测效果越好。 R^2 表达式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2} \quad (10)$$

式中: \bar{y} 为真实值均值。

2.2 岩石长短期试验与数据集准备

本文以灰砂岩作为研究对象的试验数据集对岩石常规三轴力学特性和单级松弛特性进行模拟研究。相关试验在岩石全应力多场耦合三轴试验系统上进行。试验所用岩石取自同一岩块, 主要由石英、长石等矿物胶结而成, 表面呈灰青色, 夹杂少许黑色, 颗粒较粗糙。根据国际岩石力学学会标准^[25], 试样为 $50 \text{ mm} \times 100 \text{ mm}$ 的标准圆柱形尺寸, 试样各部位遵循相关标准的精度规定。所有试样的平均质量为 451 g , 平均密度为 2.3 g/cm^3 。试验仪器及试样如图 3 所示。



图 3 岩石全应力多场耦合三轴试验系统 (左) 与灰砂岩岩石试样 (右)

Fig.3 Rock full stress multi-field coupled triaxial testing system (left) and gray sandstone rock sample (right)

考虑到试验数据有限, 通常只有固定数量的围压试验结果, 且预测岩石短长期力学特性的 LSTM 本构模型最终预测的是岩石短长期变形曲线。因此训练集和测试集的划分不能采用常见的随机比例分配方法, 应当从中间围压或偏压的试验数据中随机抽取一组作为测试集。

本文使用灰砂岩在 0、5、10、15、20、25、30、40 MPa 共 8 个围压等级下, 开展的常规三轴压缩试验所获得的相应试验数据, 即体积应变 ε_v 、偏应力 $\sigma_1 - \sigma_3$ 、环向应变 ε_3 、轴向应变 ε_1 、围压 σ_3 作为研究的数据样本集。其中环向应力 σ_3 为试验设定的围压值, σ_1 为轴压。对于试样短期力学特性研究, 本文选取 $\sigma_3=25 \text{ MPa}$ 等级下的数据样本作为验证模型结果的测试集, 测试集样本数量为 3 930 个; 其余 7 个围压等级下的数据样本作为训练集, 训练集样本数量为 26 678 个。

对于试样长期力学特性研究, 本文将 7 个试样在围压等级为 20 MPa 时进行的不同初始偏应力 $(\sigma_1 - \sigma_3)^0$, 松弛时间为 36 h 的单级松弛所获得的试验数据作为深度学习研究的数据样本集, 选取初始偏压值为 119、136、144、152、167、175 MPa 的数据样本作为训练集, 训练集样本数量为 151 446

个；选取初始偏压值为 128 MPa 的数据样本作为验证模型结果的测试集，测试集样本数量为 25 201 个。

考虑到模型的输入、输出参数间单位不同，直接进行分析会出现模型精度不够的问题，同时量纲不会导致每次迭代更新模型参数的速度下降。为了避免上述问题，本文在数据预处理阶段对总数据样本集的输入参数和输出参数进行归一化，将数据线性缩放至[0,1]区间，具体算法如下：

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (11)$$

式中： x 为输入和输出参数的真实值； x_{\max} 、 x_{\min} 为各输入和输出参数真实值中的最大值、最小值； x^* 为归一化后的数据，用于模型的训练和测试。

3 基于 LSTM 的灰砂岩短长期力学特性研究

3.1 模型构建

深度学习中的神经网络通常由输入层、隐藏层和输出层组成。LSTM 在隐藏层中具有循环连接，使其能捕捉到序列信息，如图 4 所示。其中，箭头表示网络传播方向，上标 (1) 表示第 1 层，以此类推。 x_u 表示输入层第 u 个神经元，同理 y_v 表示输出层第 v 个神经元。根据问题的复杂程度，可以调整隐藏层的层数和神经元数，输入层和输出层的神经元数分别由输入参数和输出参数个数决定。

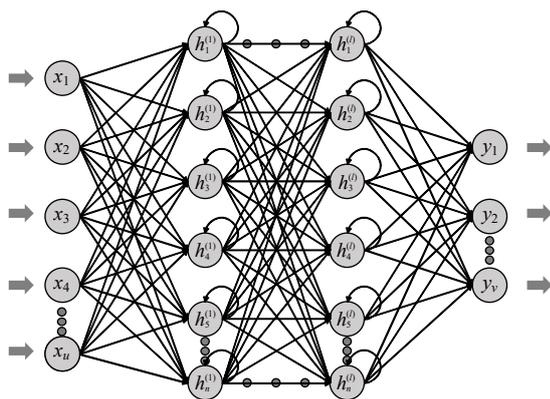


图 4 LSTM 网络拓扑结构
Fig.4 LSTM network topology

神经网络的训练过程主要包含前向传播、反向传播和迭代更新模型参数。训练初始，模型参数一般为随机值，因此通过前向传播得到的结果必然是不可靠的，而预测结果与目标值之间的误差通常由损失函数衡量。本文 LSTM 模型采用的损失函数是光滑函数均方误差 MSE (mean squared error)：

$$MSE = \frac{\sum_{i=1}^n [y - y_i]^2}{n} \quad (12)$$

MSE 越小，意味着模型预测效果越好。因此训练模型的过程可转化为损失函数最小化问题。常规神经网络一般通过优化算法使得损失函数达到极小值，从而确定对应的网络参数。优化算法通常包括梯度下降算法和进化算法。本文 LSTM 模型采用梯度下降算法中的自适应矩估计算法 (adaptive moment estimation, 简称 ADAM) 对模型进行训练。ADAM 优化器是由 Kingma 等^[26]于 2015 年提出，ADAM 通过计算梯度的一阶矩估计和二阶矩估计为不同参数设计独立的自适应性学习率，通常参数调整较少。

3.2 基于 LSTM 的短期力学特性研究

本节所用数据集是灰砂岩在常规三轴压缩试验下获得的试验数据集，涉及有体积应变 ε_v 、偏应力 $\sigma_1 - \sigma_3$ 、环向应变 ε_3 、轴向应变 ε_1 、围压 σ_3 共 5 个参数。试验按应变控制加载，且体积应变 ε_v 可由计算得到。利用 LSTM 模型确定灰砂岩短期力学特性的问题可以转化为参数为 4，输出参数为 2 的神经网络模型。为获得最优模型，本文通过随机网格搜索 (randomized search cross-validation, 即 Randomized Search CV) 算法来进行自动超参数调优选取。在超参数优化选取过程中，分别从是否加入随机失活 (Dropout) 层、每层不同神经元数、LSTM 模型层数 3 个方面设置模型结构和超参数范围，并通过 RMSE 大小来确定最优超参数组合。

3 组超参数的具体选取范围如下：第 1 组为 LSTM 模型与增加 Dropout 层的 LSTM 模型；第 2 组为 LSTM 模型每层神经元数，具体取值为 20、40、60、80、100、120；第 3 组为 LSTM 模型层数，具体取值为 1、2、3。

3.2.1 LSTM 模型与加入 Dropout 层的 LSTM 模型对比

Dropout 可以增强模型的泛化性、缓解模型过拟合，但会导致梯度下降变慢，模型的收敛速度减小。因此，最优模型的选取应当考虑对比 LSTM 模型与增加 Dropout 层的 LSTM 模型。

LSTM 模型与 LSTM+Dropout 模型测试集的损失函数随训练轮次的变化曲线如图 5 所示。两个模型相应评价指标结果见表 1。本文模型涉及的应力、应变输入参数都具有很强的时间相关性，Dropout 层的随机失活却会导致模型损失部分记忆，从而使模型评价指标 RMSE、MAE 增大， R^2 减小。

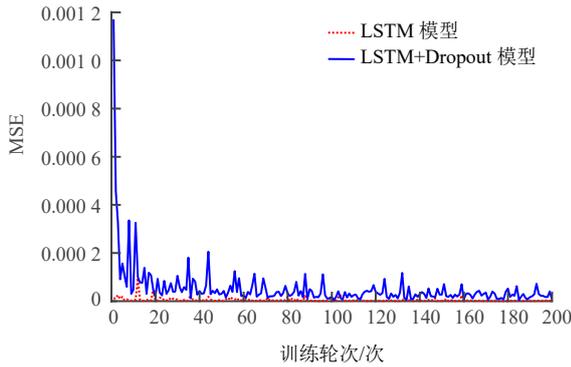


图 5 灰砂岩常规三轴压缩 LSTM 模型与 LSTM+Dropout 模型测试集均方差演化曲线

Fig.5 Mean square error evolution curves of gray sandstone conventional triaxial compression LSTM model and LSTM+Dropout model test set

表 1 灰砂岩常规三轴压缩 LSTM 模型与 LSTM+Dropout 模型测试集评价指标与训练时间对比
Table 1 Comparison of test set evaluation indicators and training time between conventional triaxial compression LSTM model and LSTM+Dropout model for gray sandstone

模型	神经元数/个	LSTM 网络层数/层	均方根误差 RMSE	平均绝对误差 MAE	决定系数 R^2	训练时间/训练轮次 $(s \cdot \text{次}^{-1})$
LSTM	80	1	0.000 747	0.000 622	0.999 9	4
LSTM+Dropout	80	1	0.002 946	0.000 622	0.999 8	4

3.2.2 不同神经元数的 LSTM 模型对比

神经网络由众多神经元通过权重连接而成。隐藏层的神经元数不定，数目越多，神经网络的非线性越显著。

本节不同神经元数的 6 个 LSTM 模型测试集的损失函数 MSE 随训练轮次的变化曲线如图 6 所示。表 2 汇集了对应的模型评价指标与训练时间。

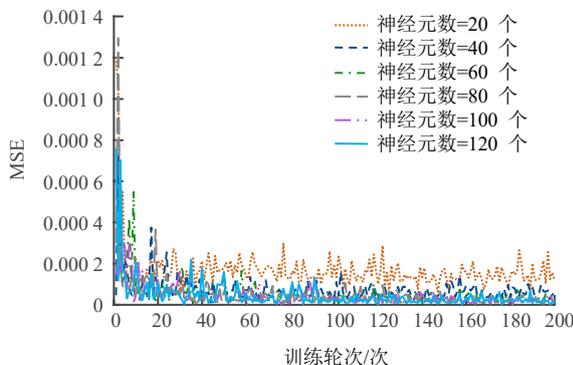


图 6 灰砂岩常规三轴压缩不同神经元数的 LSTM 模型测试集均方差演化曲线

Fig.6 Evolution curves of the mean square error of the LSTM model test set for conventional triaxial compression of gray sandstone with different numbers of neurons

表 2 灰砂岩常规三轴压缩不同神经元数的 LSTM 模型测试集评价指标与训练时间对比

Table 2 Comparison of test set evaluation index and training time of LSTM model with different neuron numbers for conventional triaxial compression of gray sandstone

神经元数/个	LSTM 网络层数/层	均方根误差 RMSE	平均绝对误差 MAE	决定系数 R^2	训练时间/训练轮次 $(s \cdot \text{次}^{-1})$
20	1	0.010 867	0.005 351	0.997 8	4
40	1	0.007 684	0.005 725	0.998 5	4
60	1	0.003 814	0.003 122	0.999 5	4
80	1	0.003 234	0.002 682	0.999 6	4
100	1	0.003 186	0.002 837	0.999 6	6
120	1	0.005 351	0.004 118	0.999 4	6

3.2.3 不同 LSTM 层数的模型对比

隐藏层在输入层和输出层中间，由大量神经元及链接所构成。增加隐藏层层数通常可以达到降低误差、提高模型精度的作用，但同时隐藏层层数的增加也使网络变得复杂化，延长模型的训练时间，并使模型出现过拟合等问题。

本节针对不同 LSTM 隐藏层层数对模型的影响进行分析研究，图 7 为 1、2、3 层 LSTM 隐藏层模型测试集的损失函数 MSE 随训练轮次的变化曲线。3 个模型的评价指标和训练时间见表 3。

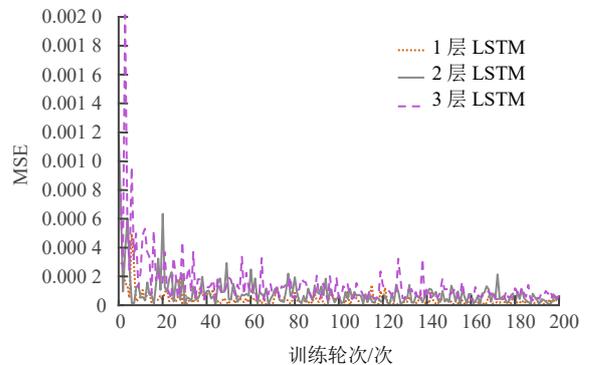


图 7 灰砂岩常规三轴压缩不同 LSTM 层数的模型测试集均方差演化曲线

Fig.7 Evolution curves of the mean square error of the model test set for conventional triaxial compression of gray sandstone with different numbers of LSTM layers

表 3 灰砂岩常规三轴压缩不同 LSTM 层数的模型测试集评价指标与训练时间对比

Table 3 Comparison of model test set evaluation indicators and training time with different numbers of LSTM layers for conventional triaxial compression of gray sandstone

LSTM 网络层数/层	神经元数/个	均方根误差 RMSE	平均绝对误差 MAE	决定系数 R^2	训练时间/训练轮次 $(s \cdot \text{次}^{-1})$
1	80	0.006 831	0.005 384	0.998 6	4
2	80	0.006 392	0.005 430	0.999 5	7
3	80	0.007 068	0.005 988	0.998 1	10

3.2.4 短期力学特性研究模型结果

通过随机网格搜索算法进行超参数的最优组合选取，可以确定不加入 Dropout 层、隐藏层神经元数为 80 个、隐藏层层数为 2 层的 LSTM 网络模型结构为最优选择，模型的 RMSE 为 0.001 890，适合研究灰砂岩常规三轴力学特性。图 8 描绘了模型对灰砂岩在围压等级为 25 MPa 下常规三轴压缩特性的预测结果，并与试验结果进行对比。

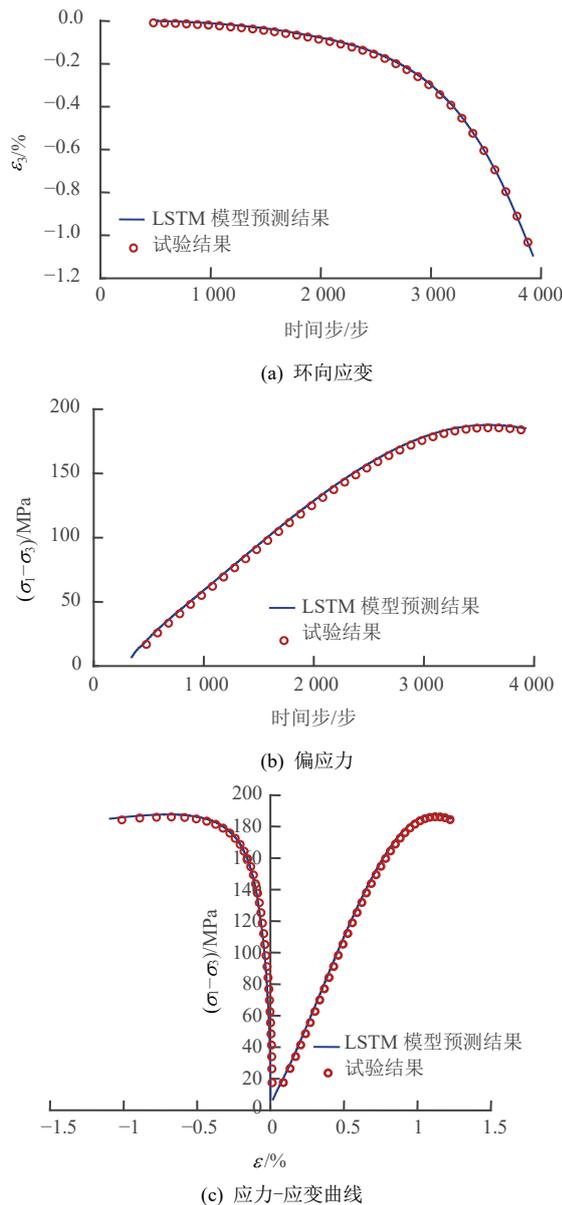


图 8 灰砂岩常规三轴力学特性 LSTM 模型预测结果与试验结果对比

Fig.8 Comparison between LSTM model prediction results and experimental results of conventional triaxial mechanical properties of gray sandstone

在一般情况下，岩石试样的三轴压缩过程可以 4 个特征应力点，即初始损伤闭合应力 σ_{cc} 、起裂应力 σ_{ci} 、裂纹损伤应力 σ_{cd} 和峰值强度 σ_c ，划分为 5

个阶段^[27]。非线性初始损伤压密闭合阶段：天然灰砂岩内部的张开性裂隙受偏应力作用开始逐渐闭合，到达初始损伤闭合应力 σ_{cc} 时，试样内部裂纹基本完全闭合；线弹性加载阶段：岩石基质发生变形，曲线呈线性上升，通过此阶段的应力-应变曲线可以获得岩样的弹性模量和泊松比变形参数，起裂应力 σ_{ci} 为该阶段的弹性极限点；裂纹稳定扩展阶段：岩样内部萌生新的微裂纹，开始产生无法恢复的塑性变形，裂纹损伤应力 σ_{cd} 为该阶段的上限点；裂纹不稳定扩展阶段：岩样内部不仅不断萌生新裂纹，这些裂纹也在不停地扩展、搭接并贯通，最终岩样达到峰值强度 σ_c ；峰后软化阶段：岩样承载力降低发生破坏，最后应力稳定于某一恒定值。

从图 8 中可以看出，通过 LSTM 搭建的网络模型能够很好地预测灰砂岩试样三轴压缩试验结果，且预测得到的应力-应变曲线满足上述岩样三轴压缩过程中 5 阶段的变化规律，能准确地反映灰砂岩的偏应力和环向应变的非线性以及应变硬化等宏观力学特性。

3.3 基于 LSTM 的长期力学特性研究

本节研究对象是灰砂岩应力松弛试验数据。试验通过应变恒定模式使试样进入松弛阶段，涉及时间 t 、体积应变 ε_v 、偏应力 $\sigma_1 - \sigma_3$ 、围压 σ_3 、孔压 p_w 、环向应变 ε_3 、轴向应变 ε_1 、温度 T 共 8 个参数。本节主要研究松弛阶段灰砂岩偏应力随时间的变化曲线，考虑到试验过程中对试验室进行了隔热处理，降低了温度对试验的影响，且 7 组松弛试验灰砂岩孔压相近，基本趋于 0，因此采用前一刻的偏应力 $\sigma_1 - \sigma_3$ 、当前时刻的轴向应变 ε_1 和时间 t 以及初始偏应力 $(\sigma_1 - \sigma_3)^0$ 作为输入，当前时刻的偏应力 $\sigma_1 - \sigma_3$ 作为输出。为获得最优模型，对比模型构建方案如表 4 所示。

表 4 灰砂岩松弛特性 LSTM 模型构建方案
Table 4 LSTM model construction plans for relaxation properties of gray sandstone

组号	序号	隐藏层构成	层数/层	神经元数/个	训练轮次/次
第 1 组	1	LSTM	1	100	200
	2	LSTM+Dropout	1	100	200
	3	LSTM+Dropout	1	20	200
	4	LSTM+Dropout	1	40	200
第 2 组	5	LSTM+Dropout	1	60	200
	6	LSTM+Dropout	1	80	200
	7	LSTM+Dropout	1	100	200
	8	LSTM+Dropout	1	120	200
第 3 组	9	LSTM+Dropout	1	100	200
	10	LSTM+Dropout	2	100	200
	11	LSTM+Dropout	3	100	200

使用表 4 中的 11 种 LSTM 模型模拟灰砂岩应力松弛力学特性, 各模型测试集的损失函数 MSE 随训练轮次的变化曲线如图 9~11 所示, 评价指标结果见表 5。

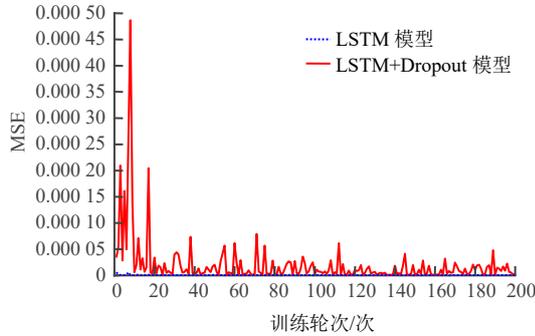


图 9 灰砂岩应力松弛 LSTM 模型与 LSTM+Dropout 模型测试集均方误差演化曲线

Fig.9 Mean square error evolution curves of gray sandstone stress relaxation LSTM model and LSTM+Dropout model test set

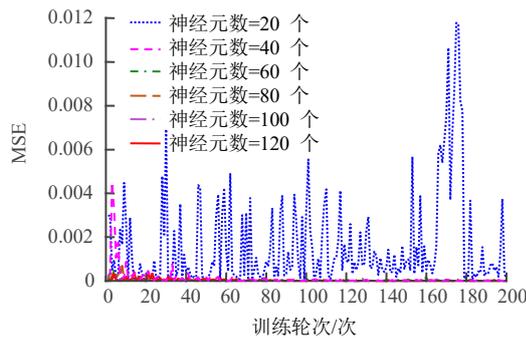


图 10 灰砂岩应力松弛不同神经元数的 LSTM 模型测试集均方误差演化曲线

Fig.10 Evolution curves of mean square error of LSTM model test set for stress relaxation of gray sandstone with different numbers of neurons

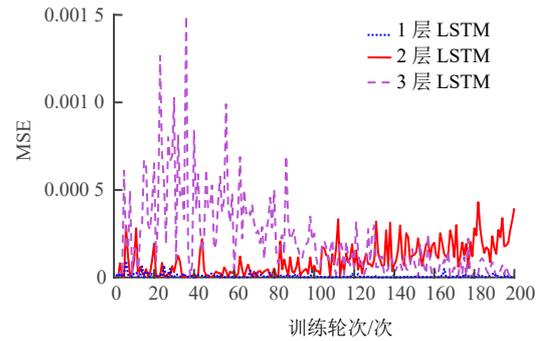


图 11 灰砂岩应力松弛不同 LSTM 层数的模型测试集均方误差演化曲线

Fig.11 Model test set mean square error evolution curves of gray sandstone stress relaxation with different numbers of LSTM layers

通过随机网格搜索算法进行超参数的最优组合选取, 可以确定不加 Dropout 层、隐藏层数为 1 层、神经元数为 100 个的 LSTM 模型结构为最优选择, 模型的 RMSE 为 0.000 656, 适合研究灰砂岩应力松弛特性。为消除设备和温度造成的试验误差, 选用拟合后的试验结果对选用模型进行重新训练, RMSE、MAE、 R^2 分别为 0.000 743、0.000 725、0.995 700。图 12 为模型对灰砂岩在围压为 20 MPa、初始偏应力为 128 MPa 时的单级松弛预测结果。与拟合结果进行对比。可以发现, 模型预测结果与拟合结果高度一致, 说明 LSTM 模型可以反映灰砂岩的应力松弛特性。

从训练集曲线形态可以看出, 不同初始偏应力的灰砂岩试样单级松弛变化趋势类似, 松弛的过程主要分为 3 个阶段: ①快速松弛阶段。岩样在进入松弛后, 偏应力迅速降低, 这个过程持续 30 min 左右。②减速松弛阶段。偏应力在此阶段仍持续下降,

表 5 灰砂岩应力松弛模型测试集评价指标与训练时间对比

Table 5 Comparison of test set evaluation indicators and training time of gray sandstone stress relaxation model

序号	模型	神经元数 / 个	LSTM 网络层数 / 层	均方根误差 RMSE	平均绝对误差 MAE	决定系数 R^2	训练时间/训练轮次/(s · 次 ⁻¹)
1	LSTM	100	1	0.000 656	0.000 477	0.997 3	30
2	LSTM+Dropout	100	1	0.001 740	0.001 495	0.981 3	31
3	LSTM+Dropout	20	1	0.010 528	0.007 447	0.313 6	31
4	LSTM+Dropout	40	1	0.003 168	0.002 637	0.937 9	31
5	LSTM+Dropout	60	1	0.005 377	0.005 260	0.820 9	31
6	LSTM+Dropout	80	1	0.001 870	0.001 540	0.978 3	31
7	LSTM+Dropout	100	1	0.001 656	0.001 425	0.983 0	19
8	LSTM+Dropout	120	1	0.005 153	0.005 078	0.835 6	52
9	LSTM+Dropout	100	1	0.001 737	0.001 487	0.981 3	49
10	LSTM+Dropout	100	2	0.019 886	0.019 812	-1.449 1	90
11	LSTM+Dropout	100	3	0.005 004	0.004 881	0.844 9	134

但下降速率逐渐放缓, 该过程所需时间一般为 12 h。③稳态松弛阶段。进入稳态阶段后, 偏应力趋于恒

定值, 应力以接近于 0 的松弛速率衰减。若试验时间足够长, 且无其他外力影响, 应力松弛量将恒定

不变。图 12 的曲线反映了通过 LSTM 搭建的网络模型预测得到的灰砂岩试样应力松弛时偏应力与时间的关系，可以发现预测得到的曲线形态满足岩样单级松弛过程中 3 阶段的变化规律。

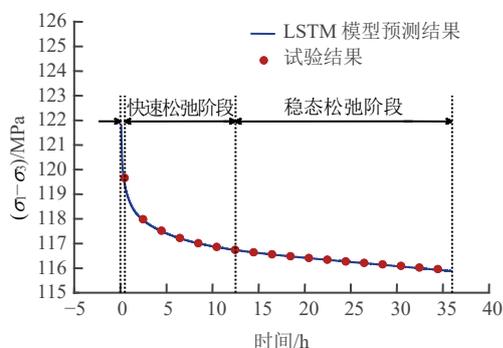


图 12 灰砂岩应力松弛力学特性 LSTM 模型预测结果与拟合结果对比

Fig.12 Comparison between LSTM model prediction results and fitting results of stress relaxation mechanical properties of gray sandstone

3.4 基于注意力机制的 LSTM 的灰砂岩短长期力学特性研究

引入注意力机制可以在一定程度上提高 LSTM 模型的适用范围和性能表现。此外，注意力机制还可以帮助解决长序列训练过程中的梯度消失问题，从而提高模型的训练效率和泛化能力。基于注意力机制的 LSTM 模型由输入层、LSMT 隐藏层、注意力层、全连接层以及输出层组成。根据注意力层所处位置的不同，其捕捉的关键信息有所不同。

3.4.1 基于注意力机制的短期力学特性研究

本节对灰砂岩常规三轴力学特性采用的输入输出参数与 3.2 节基于 LSTM 的模型相同。利用时间注意力结构和特征注意力结构对不加入 Dropout 层、隐藏层神经元数为 80 个、隐藏层层数为 2 层的 LSTM 模型进行优化。4 种模型分别为：LSTM-Time 模型、LSTM-Feature 模型、Time-LSTM 模型、Feature-LSTM 模型。

分别使用 4 种模型对围压为 25 MPa 下的灰砂岩常规三轴力学特性进行预测。经过 100 轮训练，可以发现，基于时间注意力机制的 LSTM 模型和基于特征注意力机制的 LSTM 模型对灰砂岩常规三轴力学特性预测精度极高。与传统的 LSTM 模型预测效果对比，单次训练时间较不加入注意力机制的 LSTM 模型增加了 1 s，模型预测精度无明显变化，决定系数 R^2 均达 0.99 以上。

对 5 组 LSTM 模型测试集的测试过程进行调整，除第 1 组序列长度输入数据仍采用真实试验结

果外，其他序列长度输入数据仅保留按应变加载时试验。已知的当前时刻的轴向应变 ε_1^t 和环向应力 σ_3^t ，前一时刻的偏应力 $(\sigma_1 - \sigma_3)^{t-1}$ 和环向应变 ε_3^{t-1} 则替换成上一组的预测结果。可以发现，与传统的 LSTM 模型调整后的预测结果精度较低，且受模型训练过程影响评价指标值不稳定。相较而言，增加注意力机制的 LSTM 模型对于未知变形的预测更为准确。

3.4.2 基于注意力机制的长期力学特性研究

本节对灰砂岩应力松弛力学特性采用输入、输出参数与 3.3 节中基于 LSTM 的模型相同。利用时间注意力结构和特征注意力结构不加入 Dropout 层、隐藏层神经元数为 1 000、隐藏层层数为 1 层的 LSTM 模型进行优化。4 种模型分别为：LSTM-Time 模型、LSTM-Feature 模型、Time-LSTM 模型、Feature-LSTM 模型。

使用前述的 4 个模型对初始偏应力为 128 MPa 的灰砂岩应力松弛力学特性进行预测。经过 50 轮训练，可以发现，在 LSTM 层前加入时间注意力和特征注意力两个模型决定系数 R^2 达到了 0.97 以上，拟合效果较好。与传统的 LSTM 模型预测效果对比，加入注意力机制增加了对应模型训练时间，使用注意力机制的模型较未使用注意力的模型预测精度没有明显提高，而在 LSTM 层后增加注意力机制导致模型训练过程更不稳定，预测效果也出现大幅降低。

4 有限元分析中 LSTM 深度学习模型的实现方法

第 3 章节基于 LSTM 模型对灰砂岩的常规三轴力学特性和单级应力松弛特性进行研究预测，有效验证了深度学习方法对灰砂岩长短时力学状态的预测能力。本节运用有限元程序，将通过深度学习方法确定的网络模型编写为本构程序，完成深度学习模型的数值实现。通过将模型数值结果与试验结果进行对比，验证网络模型准确性的同时，进一步彰显深度学习方法在建立本构模型时的优势。

4.1 本构模型数值实现步骤

有限元方法是一种数值分析方法，主要用于求解偏微分方程。它将连续的物理问题转化为离散的数学问题，然后使用数值方法求解。本节将利用 Julia 程序语言编制的，可用于结构计算的三维有限元框架实现 LSTM 深度学习网络。在有限元程序中，小变形固体力学初边值问题的数学表达式通过应用变分法或最小势能原理，可得到原问题的弱形式。再应用伽辽金法，进行空间离散，用单元集合代替原物理空间，原问题的弱形式被离散成单元上的和。

单元内任一点的位移可用单元结点位移插值表示。由此,无限自由度的偏微分方程初边值问题,转化成有限自由度的问题,利用诸如 Newton-Raphson 法等数值解法可以求得近似解。而插入的训练模型在有限元程序中起着类似本构规律的作用。该程序的关键在于使用当前应力 σ_i 、应变 ε_i 以及规定应变增量 $d\varepsilon$ 来更新应力张量 $\sigma^{t+\Delta t}$ 和 Jacobian 矩阵 $\partial\Delta\sigma/\partial\Delta\varepsilon$ 。

LSTM 模型的输入是序列数据,即输入是由之前时间步的应力、应变结果和当前时间步的应力、应变结果共同组成,但考虑到有限元程序在求解某一时间步的应力和 Jacobian 矩阵时仅向本构程序提供前一时刻的应力 σ 和应变 ε 以及当前时刻的应变增量 $d\varepsilon$,并不构成模型所需的序列数据。因此,为获得 LSTM 模型所需的初始序列数据完成模型的数值实现。本文将瞬时本构程序求解过程分为两部分,分别是各向同性线弹性本构模型和 LSTM 网络模型。在初始阶段,本构程序首先执行各向同性线弹性本构模型,并保存当前时刻的结果,直至获得所需长度的序列数据,再进入 LSTM 模型中求解。为了存放之前时刻的应力应变结果,从而构成 LSTM 模型所需的序列数据输入,本文设定状态变量 STATEV 进行数据存储,STATEV 存储空间为 $N\times L+1$ (N 为输入参数数量, L 为输入序列长度)。将前一时刻的偏应力和环向应变以及当前时刻的轴向应变和环向应力按时间顺序依次存放在 STATEV 中,当 STATEV 中 $1\sim N\times L$ 号位存放完毕,即获得一组完整的序列数据。为了确保序列数据存放完毕,设定 STATEV 中第 $N\times L+1$ 号位为计数位,用于判断本构程序执行次数。为了确保序列数据存放瞬时本构程序执行过程以及 STATEV 具体存放和更新流程如图 13 所示。

4.2 基于深度学习的灰砂岩瞬时本构模型

灰砂岩瞬时本构模型采用由特征注意力机制、两层 LSTM 隐藏层和一个全连接层组成的网络模型实现。若模型特征值和目标值的数量分别为 j 和 k ,特征注意力机制层的神经元个数为 j ,第 1、2LSTM 隐藏层分别有 p 、 q 个神经元,全连接层神经元有 k 个,则模型内部参数数量为 $(j^2+j)+4(p^2+jp+p)+4(q^2+pq+q)+(qk+k)$ 。模型的特征值为 4、目标值为 2,注意力机制层的神经元数为 4,两层 LSTM 层神经元数均为 80,全连接层神经元数为 2,因此模型内部参数共有 78 902 个。

训练后的 Feature-LSTM 模型可以通过各层的权重和偏置参数表示应力-应变关系,各层具体求解

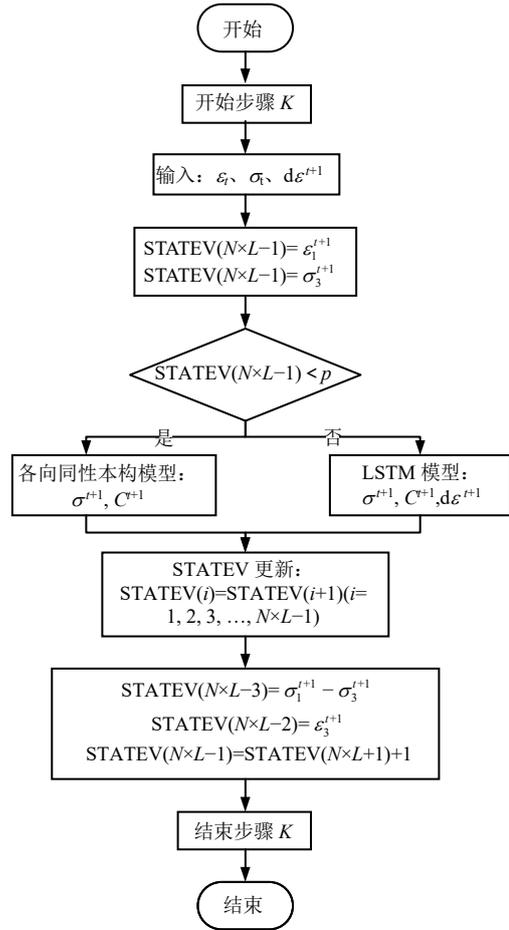


图 13 LSTM 网络本构程序执行过程流程图

Fig.13 Flowchart of implementation of LSTM network in FEM program

过程如下:

(1) 特征注意力层由全连接层和 Softmax 函数组成

$$\left. \begin{aligned} X_i &= W^{(1)}x_i + b^{(1)} \\ X'_i &= \frac{e^{X_i}}{\sum_i e^{X_i}} \\ X''_i &= x_i \odot X'_i \end{aligned} \right\} \quad (13)$$

式中: X'_i 、 X''_i 为特征注意力机制层的输出; 当前时刻的输入 x_i 由多个时间步的应力分量和应变分量组成。

(2) 第 1 层 LSTM 层

$$\left. \begin{aligned} f_i^{(1)} &= \sigma(W_f^{(1)}h_{t-1}^{(1)} + U_f^{(1)}X''_i + b_f^{(1)}) \\ i_i^{(1)} &= \sigma(W_i^{(1)}h_{t-1}^{(1)} + U_i^{(1)}X''_i + b_i^{(1)}) \\ \tilde{C}_i^{(1)} &= \tanh(W_C^{(1)}h_{t-1}^{(1)} + U_C^{(1)}X''_i + b_C^{(1)}) \\ C_i^{(1)} &= C_{i-1}^{(1)} \odot f_i^{(1)} + i_i^{(1)} \odot \tilde{C}_i^{(1)} \\ o_i^{(1)} &= \sigma(W_o^{(1)}h_{t-1}^{(1)} + U_o^{(1)}X''_i + b_o^{(1)}) \\ h_i^{(1)} &= o_i^{(1)} \odot \tanh(C_i^{(1)}) \end{aligned} \right\} \quad (14)$$

式中： $h_t^{(1)}$ 为第 1 层 LSTM 隐藏层的输出； $h_t^{(2)}$ 为第 2 层 LSTM 隐藏层的输出。

(3) 第 2 层 LSTM 层

$$\left. \begin{aligned} f_t^{(2)} &= \sigma(W_f^{(2)}h_{t-1}^{(2)} + U_f^{(2)}h_t^{(1)} + b_f^{(2)}) \\ i_t^{(2)} &= \sigma(W_i^{(2)}h_{t-1}^{(2)} + U_i^{(2)}h_t^{(1)} + b_i^{(2)}) \\ \tilde{C}_t^{(2)} &= \tanh(W_C^{(1)}h_{t-1}^{(1)} + U_C^{(1)}h_t^{(1)} + b_c^{(2)}) \\ C_t^{(2)} &= C_{t-1}^{(2)} \odot f_t^{(2)} + i_t^{(2)} \odot \tilde{C}_t^{(2)} \\ o_t^{(2)} &= \sigma(W_o^{(2)}h_{t-1}^{(2)} + U_o^{(2)}h_t^{(1)} + b_o^{(2)}) \\ h_t^{(2)} &= o_t^{(2)} \odot \tanh(C_t^{(2)}) \end{aligned} \right\} \quad (15)$$

(4) 加入 Sigmoid 激活函数的全连接层：

$$\left. \begin{aligned} s &= W^{(2)}h_t^{(2)} + b^{(2)} \\ s' &= \frac{1}{1 + e^{-s}} \end{aligned} \right\} \quad (16)$$

式中： s 为全连接层的输出； s' 为全连接层的输出，由应力分量和应变分量组成，具体为 $s' = [(\sigma_1 - \sigma_3)', \varepsilon_3']$ 。

通过 Python 程序存储并提取训练后的网络内部参数在本构程序中实现上述模型。由于模型含有大量参数，显示表达求解过程涉及过多参数，因此在本构程序中采用循环的方式实现各网络层的计算。将上述本构程序嵌入有限元框架中，按应变加载，总应变设为 0.012，模拟围压为 25 MPa 下灰砂岩常规三轴压缩力学特性，加载步数分别为 500、1 000、2 500、5 000 步，每一加载步的应变增量分别为 2.4×10^{-5} 、 1.2×10^{-5} 、 4.8×10^{-6} 、 2.4×10^{-6} 。从图 14 可以看出，4 组不同加载步的模拟曲线有一定不同，但在到达峰值应力前均与试验数据有良好的—致性，且随着加载步数的增大，模拟曲线与试验数据间的契合程度越高。

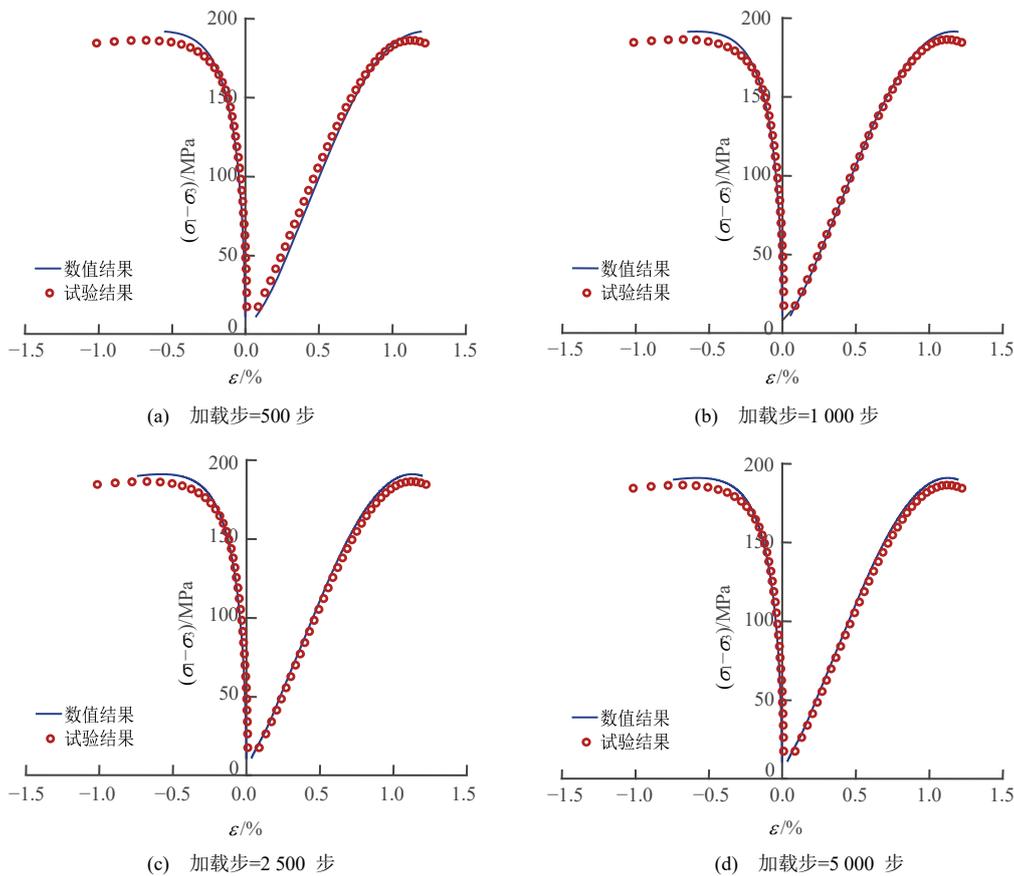


图 14 基于 Feature-LSTM-FEM 程序的灰砂岩瞬时模拟曲线与试验数据对比

Fig.14 Comparison of instantaneous simulation curves and experimental data of gray sandstone based on the Feature-LSTM-FEM program

4.3 基于深度学习的灰砂岩时效本构模型

本节对灰砂岩时效本构模型采用一个 LSTM 层、一个全连接层以及一个 Relu 层组成的 LSTM 网络模型实现。模型的特征值为 4、目标值为 1，隐藏

层神经元数为 100 个，全连接层神经元数为 1 个，因此模型内部参数共有 42 101 个。LSTM 模型的输入由时间、应力以及应变组成。

利用训练得到的模型参数可以将灰砂岩偏应力

随时间的变化规律表示出来,从而获得基于 LSTM 网络模型的灰砂岩时效本构模型,具体如下:

(1) 第 1 层 LSTM 层

$$\left. \begin{aligned} f_t^{(1)} &= \sigma(W_f^{(1)}h_{t-1}^{(1)} + U_f^{(1)}x_t + b_f^{(1)}) \\ i_t^{(1)} &= \sigma(W_i^{(1)}h_{t-1}^{(1)} + U_i^{(1)}x_t + b_i^{(1)}) \\ \tilde{C}_t^{(1)} &= \tanh(W_c^{(1)}h_{t-1}^{(1)} + U_c^{(1)}x_t + b_c^{(1)}) \\ C_t^{(1)} &= C_{t-1}^{(1)} \odot f_t^{(1)} + i_t^{(1)} \odot \tilde{C}_t^{(1)} \\ o_t^{(1)} &= \sigma(W_o^{(1)}h_{t-1}^{(1)} + U_o^{(1)}x_t + b_o^{(1)}) \\ h_t^{(1)} &= o_t^{(1)} \odot \tanh(C_t^{(1)}) \end{aligned} \right\} \quad (17)$$

(2) 全连接层

$$s = Wh_t^{(1)} + b \quad (18)$$

(3) Relu 激活函数层

$$f(s) = \max(0, s) \quad (19)$$

式中: $f(s)$ 为激活层的输出,表示当前时刻的偏应力。

将本构程序嵌入有限元程序中,先按应变加载,加载到恒定初始轴向应变后,模拟围压为 20 MPa、初始偏应力为 128 MPa 的灰砂岩应力松弛力学特性,松弛时间设为 36 h,时间步数分别为 900、1 800、3 600、9 000 步,对应时间增量分别为 0.040、0.020、0.010、0.004 h。从图 15 可以确定,4 组不同时间步的模拟曲线与拟合数据均有良好的一致性,且随着加载步数的减小,模拟曲线与拟合数据间的契合程度也很高。

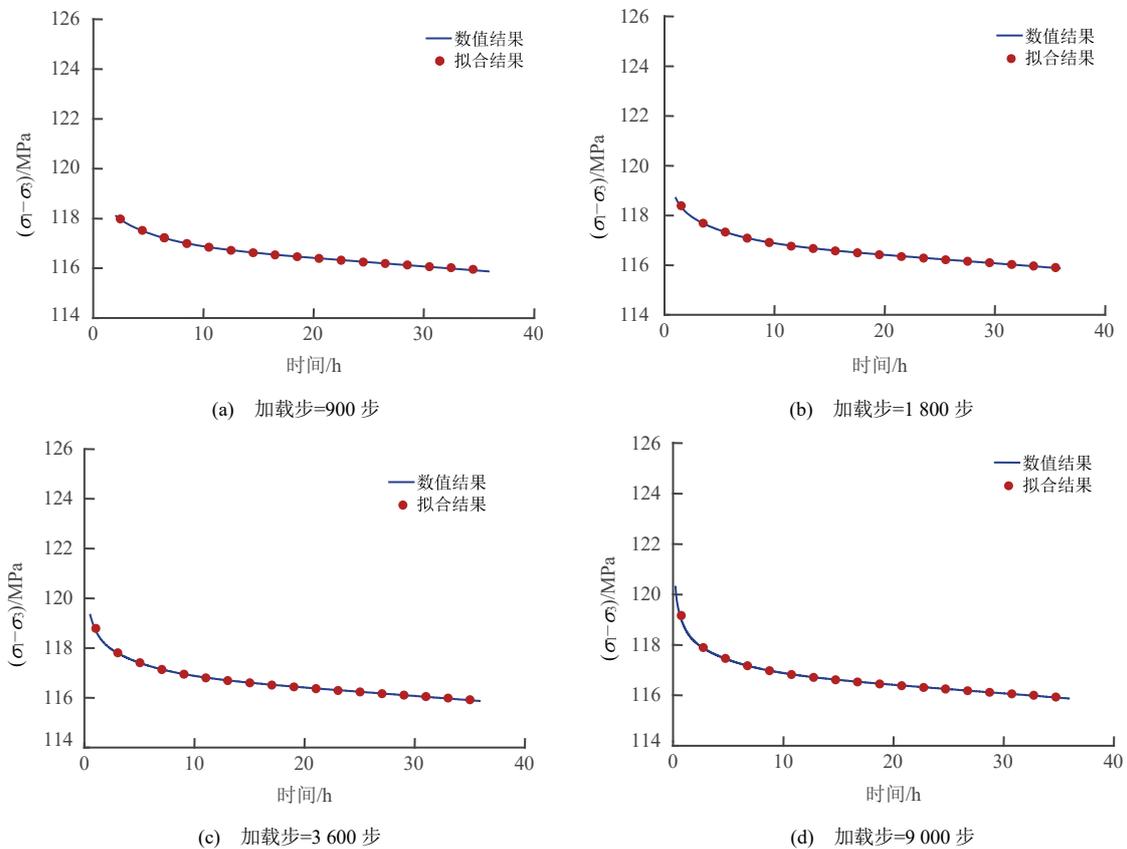


图 15 基于 LSTM-FEM 程序的灰砂岩松弛模拟曲线与拟合数据对比

Fig.15 Comparison of relaxation simulation curves of gray sandstone based on LSTM-FEM program and fitted data

5 结论

本研究主要采用 LSTM 模型对灰砂岩的常规三轴压缩特性和单级应力松弛特性进行了研究。通过将试验数据划分为训练集和测试集,构建 3 组对比模型,从而确定 2 组模型结构。获取模型后,对短期力学特性 LSTM 网络模型预测结果进行分析,主要分析内容有应力-应变曲线特征预测,并分析长期

力学特性 LSTM 网络模型。

研究表明, LSTM 预测强度基本符合两种强度准则,并满足峰值强度与围压正相关;预测弹性模量与拟合曲线结果一致。相较于传统方法, LSTM 模型在不引入任何塑性参数或本构规律的情况下准确预测得到灰砂岩的力学特性。对 LSTM 模型应力松弛预测结果进行分析,发现预测松弛曲线均满足非完全衰减特性,3 个时刻的应力松弛量较曲线虽

有一定偏差, 但整体满足变化规律。

本研究同时采用注意力机制对 LSTM 模型进行改进, 并对灰砂岩的常规三轴压缩特性和单级应力松弛特性进行了研究。考虑注意力机制的位置不同会导致捕捉的关键信息有所不同, 构建 4 组改进 LSTM 模型, 分别是: LSTM-Time 模型、LSTM-Feature 模型、Time-LSTM 模型和 Feature-LSTM 模型。获取模型后, 与第 3 节的 LSTM 模型预测结果进行对比。同时考虑到传统 LSTM 模型预测过程对未知变形下材料力学特性准确性判断的不足, 将 5 组 LSTM 模型和改进 LSTM 模型测试集的预测过程进行调整。

与 LSTM 模型预测结果进行对比, 引入注意力机制的改进 LSTM 模型对灰砂岩常规三轴压缩特性预测精度很高, 同样可以准确地描述灰砂岩的常规三轴压缩特性。对于灰砂岩长期力学特性研究, 采用拟合后的数据结果对模型进行重新训练并调整测试过程。通过对比, 表明调整后的 LSTM 模型可以对灰砂岩松弛力学特性进行有效预测。

参 考 文 献

- [1] MIURA K, OKUI Y, HORII H. Micromechanics-based prediction of creep failure of hard rock for long-term safety of high-level radioactive waste disposal system[J]. **Mechanics of Materials**, 2003, 35(3/6): 587-601.
- [2] ZHOU H, JIA Y, SHAO J F. A unified elastic-plastic and viscoplastic damage model for quasi-brittle rocks[J]. **International Journal of Rock Mechanics Mining Sciences**, 2008, 45(8): 1237-1251.
- [3] 陈星, 李建林, 邓华锋, 等. 卸荷蠕变条件下软硬相接岩层非协调变形规律研究[J]. **岩土力学**, 2023, 44(1): 303-316.
CHEN Xing, LI Jian-lin, DENG Hua-feng, et al. Uncoordinated deformation of soft and hard interconnecting strata under unloading creep conditions[J]. **Rock and Soil Mechanics**, 2023, 44(1): 303-316.
- [4] 齐亚静, 姜清辉, 王志俭, 等. 改进西原模型的三维蠕变本构方程及其参数辨识[J]. **岩石力学与工程学报**, 2012, 31(2): 347-355.
QI Ya-jing, JIANG Qing-hui, WANG Zhi-jian, et al. 3D creep constitutive equation of modified Nishihara model and its parameters identification[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2012, 31(2): 347-355.
- [5] 张向东, 李永靖, 张树光, 等. 软岩蠕变理论及其工程应用[J]. **岩石力学与工程学报**, 2004, 23(10): 1635-1635.
ZHANG Xiang-dong, LI Yong-jing, ZHANG Shu-guang, et al. Creep theory of soft rock and its engineering application[J]. **Chinese Journal of Rock Mechanics and Engineering**, 23(10): 1635-1639.
- [6] YANG C, DAEMEN J, YIN J H. Experimental investigation of creep behavior of salt rock[J]. **International Journal of Rock Mechanics and Mining Sciences**, 1999, 36(2): 233-242.
- [7] LI W Q, LI X D, HAN B, et al. Recognition of creep model of layer composite rock mass and its application[J]. **Journal of Central South University of Technology**, 2007, 14(Suppl. 1): 329-331.
- [8] ZHOU H W, WANG C P, HAN B B, et al. A creep constitutive model for salt rock based on fractional derivatives[J]. **International Journal of Rock Mechanics and Mining Sciences**, 2011, 48(1): 116-121.
- [9] 徐卫亚, 杨圣奇, 褚卫江. 岩石非线性黏弹塑性流变模型(海海模型)及其应用[J]. **岩石力学与工程学报**, 2006, 25(3): 433-447.
XU Wei-ya, YANG Sheng-qi, CHU Wei-jiang. Nonlinear viscoelasto-plastic rheological model (Hohai model) of rock and its engineering application[J]. **Chinese Journal of Rock Mechanics and Engineering**, 2006, 25(3): 433-447.
- [10] KACHANOV L M. Rupture time under creep conditions[J]. **International Journal of Fracture**, 1999, 97(1): 11-18.
- [11] DRAGON A, MROZ Z. A continuum model for plastic-brittle behaviour of rock and concrete[J]. **International Journal of Engineering Science**, 1979, 17(2): 121-137.
- [12] SHAO J F, JIA Y, KONDO D, et al. A coupled elastoplastic damage model for semi-brittle materials and extension to unsaturated conditions[J]. **Mechanics of Materials**, 2006, 38(3): 218-232.
- [13] ZHU Q, SHAO J F, MAINGUY M. A micromechanics-based elastoplastic damage model for granular materials at low confining pressure[J]. **International Journal of Plasticity**, 2010, 26(4): 586-602.
- [14] ZHU Q Z, ZHOU C B, SHAO J F, et al. A discrete thermodynamic approach for anisotropic plastic-damage

- modeling of cohesive-frictional geomaterials[J]. **International Journal for Numerical and Analytical Method in Geomechanics**, 2009, 34(12): 1250-1270.
- [15] 李守巨, 刘迎曦, 王登刚, 等. 云峰大坝弹性参数识别的神经网络方法[J]. 水利水电技术, 2000, 31(8): 51-54. LI Shou-ju, LIU Ying-xi, WANG Deng-gang, et al. Elastic parameters identification of concrete dam in the Yunfeng projects with neural networks[J]. **Water Resources and Hydropower Engineering**, 2000, 31(8): 51-54.
- [16] 陈昌彦, 王思敬, 沈小克. 边坡岩体稳定性的人工神经网络预测模型[J]. 岩土工程学报, 2001, 23(2): 157-161. CHEN Chang-yan, WANG Si-jing, SHEN Xiao-ke. Predicting models to estimate stability of rock slope based on artificial neural network[J]. **Chinese Journal of Geotechnical Engineering**, 2001, 23(2): 157-161.
- [17] 谭云亮, 王春秋. 岩石本构关系的径向基函数神经网络快速逼近模型[J]. 岩土工程学报, 2001, 23(1): 14-17. TAN Yun-liang, WANG Chun-qiu. A fast approaching model for rock constitutive equation by radial basis function neural network[J]. **Chinese Journal of Geotechnical Engineering**, 2001, 23(1): 14-17.
- [18] ZHANG N, SHEN S L, ZHOU A, et al. Application of LSTM approach for modelling stress-strain behaviour of soil[J]. **Applied Soft Computing**, 2021, 100: 106959.
- [19] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. **Neural Computation**, 1997, 9(8): 1735-1780.
- [20] HUANG F, LI X, YUAN C, et al. Attention-emotion-enhanced convolutional LSTM for sentiment analysis[J]. **IEEE Transactions on Neural Networks and Learning Systems**, 2022, 33(9): 4332-4345.
- [21] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). [S. l.]: [s. n.], 2016: 207-212.
- [22] ZHANG P, YIN Z, JIN Y, et al. An AI - based model for describing cyclic characteristics of granular materials[J]. **International Journal for Numerical and Analytical Methods in Geomechanics**, 2020, 44(9): 1315-1335.
- [23] KARAPIPERIS K, STAINIER L, ORTIZ M, et al. Data-driven multiscale modeling in mechanics[J]. **Journal of the Mechanics and Physics of Solids**, 2021, 147: 104239.
- [24] QU T, DI S, FENG Y T, et al. Towards data-driven constitutive modelling for granular materials via micromechanics-informed deep learning[J]. **International Journal of Plasticity**, 2021, 144: 103046.
- [25] BROWN E T, Rock characterization, testing, and monitoring: ISRM suggested methods[C]//International Society for Rock Mechanics Commission on Testing Methods. [S. l.]: Pergamon Press, 1981.
- [26] KINGMA D P, BA J. ADAM: a method for stochastic optimization[C]// International Conference on Learning Representations (ICLR). [S. l.]: [s. n.], 2015.
- [27] ERARSLAN N. 脆性岩石试样断裂过程区及宏观疲劳裂纹行为研究[J]. 岩土力学, 2023, 44(7): 2041-2049. ERARSLAN N. Investigation of the fracture process zone and behavior of the macro-scale fatigue cracks in brittle rock specimens[J]. **Rock and Soil Mechanics**, 2023, 44(7): 2041-2049.