

面向冷弯型钢构件弹性屈曲临界荷载预测的 BP神经网络训练算法比选*

戴宜凌, 王少快, 尹凌峰

(东南大学 土木工程学院, 南京 211189)

摘要: 弹性屈曲临界荷载是准确评价冷弯型钢构件承载力的重要指标. 利用人工神经网络 (artificial neural networks, ANNs) 模型对冷弯 C 型截面轴压构件的屈曲临界荷载进行了预测, 将影响屈曲的几何参数和有限条法所得的计算结果作为数据集, 对神经网络模型进行了训练、验证和测试. 基于最优化理论, 采用 6 种不同的优化算法进行了模型的训练, 并比较了不同算法的网络模型性能. 通过随机网格搜索确定最优超参数, 使用 3 种统计参数来评估训练后的人工神经网络的性能, 以得到最适合预测屈曲临界荷载的神经网络模型. 结果表明: Levenberg-Marquardt (L-M) 算法在非线性最小二乘问题上相较于其他算法具有更高的准确性, 多次训练后, L-M 算法使模型预测误差非常小, 而其他算法在准确度上不及 L-M 算法.

关键词: BP神经网络; 最优化理论; 弹性屈曲临界荷载; 冷弯型钢; 非线性最小二乘

中图分类号: O31 **文献标志码:** A **DOI:** 10.21656/1000-0887.450050

Evaluation of BP Neural Network Algorithms for Predicting Elastic Buckling Loads on Cold-Formed Steel Components

DAI Yiling, WANG Shaokuai, YIN Lingfeng

(School of Civil Engineering, Southeast University, Nanjing 211189, P.R.China)

Abstract: The elastic buckling critical load is a crucial indicator for accurately assessing the load-bearing capacity of cold-formed steel components. The artificial neural networks were used to predict the buckling loads on cold-formed flanged steel columns, with geometric parameters and finite strip method results as the dataset. Six optimization algorithms based on the optimization theory were applied to train the networks, with their performances compared. Optimal hyperparameters were determined through random grid search. Three statistical parameters were used to evaluate the networks' post-training performances. The Levenberg-Marquardt (L-M) algorithm demonstrates higher accuracy in nonlinear least squares problems, significantly reducing prediction errors after multiple trainings, and outperforming other algorithms.

Key words: BP neural network; optimization theory; elastic buckling critical load; cold-formed steel; nonlinear least squares

* 收稿日期: 2024-02-22; 修订日期: 2024-04-23

基金项目: 国家自然科学基金(52278150)

作者简介: 戴宜凌(2000—), 女, 硕士 (E-mail: daiyiling1022@163.com);

尹凌峰(1974—), 男, 教授, 博士, 博士生导师 (通讯作者. E-mail: eking@seu.edu.cn).

引用格式: 戴宜凌, 王少快, 尹凌峰. 面向冷弯型钢构件弹性屈曲临界荷载预测的 BP 神经网络训练算法比选[J]. 应用数学和力学, 2025, 46(2): 129-141.

0 引言

现阶段,冷弯型钢结构在建筑业中已经扮演起了重要角色,其具有明显的优势,例如高强度自重比和刚度自重比,易于处理和运输,能够生产各种截面形状的灵活制造工艺,以及可回收性好,同时也为装配式建筑的发展提供了条件。

冷弯型钢构件通常为薄壁构件,截面形式多为开口式,使得构件屈曲行为复杂,在受压时容易出现局部屈曲、畸变屈曲或整体屈曲,这些屈曲模式在利用 Euler 公式进行屈曲计算时并未考虑。当前,直接确定构件屈曲承载力的直接强度法(direct strength method, DSM)在冷弯薄壁型钢结构领域的应用越来越广泛。与有效宽度法^[1]不同的是,直接强度法不需要计算构件的有效截面特性,而是直接采用全截面特性,对构件材料强度进行折减^[2-4]。直接强度法的基本设计思想为:首先计算各屈曲模式对应的弹性屈曲荷载,在此基础上计算出每种屈曲模式对应的承载力,将其中的最小值作为构件的极限承载力。在计算截面弹性局部屈曲临界荷载和弹性畸变屈曲临界荷载时,通常可由数值解确定,如有限条法(finite strip method, FSM)^[5], Li 和 Schafer 等^[6-7]将广义梁理论(general beam theory, GBT)的相关理论应用于有限条法,提出了约束有限条法(constrained finite strip method, CFSM),能够对构件进行特征值屈曲分析和屈曲模式分解,由此,可以得到截面弹性局部屈曲和弹性畸变屈曲临界荷载^[4,8-9]。然而,需要注意的是,基于有限条法得到弹性屈曲临界荷载的流程较为繁琐,计算效率较低,对于不同的构件截面需要建立对应的有限条模型,划分网格以及材料参数设置等,得到屈曲荷载因子与半波长度曲线,进而得到屈曲临界荷载,该方法非常耗时。冷弯型钢构件承载力的确定基于屈曲临界荷载的确定,因此如何快速、准确地得到冷弯型钢构件的弹性屈曲临界荷载是一个亟待解决的问题。

随着机器学习领域,特别是人工神经网络(ANN)的最新发展,预测构件的屈曲行为变成可能,与传统的分析方法相比有诸多优点^[10-14]。人工神经网络是模拟生物有机体学习机制的机器学习技术分支,类似于人脑的结构,包含许多神经元,能够在没有特定解决方案的给定数据集中准确建立输入与输出之间的关系^[15]。Kaveh 等^[16-19]是在结构工程领域应用人工神经网络先驱者,利用神经网络在结构领域的预测能力进行了多项研究。许多研究人员利用人工神经网络预测了土木结构的屈曲行为, Mojtabaei 等^[20]对 4 种不同截面的钢构件的轴压屈曲临界荷载和受弯屈曲临界荷载进行了人工神经网络预测。Waszczyszyn 等^[21]采用反向传播神经网络(back propagation neural network, BPNN),预测了轴压下圆柱壳的屈曲荷载,其中弹性反向传播学习算法被用于基于壳体样本缺陷数据的屈曲荷载因子预测。Markopoulos 等^[22]则利用 Levenberg-Marquardt (LM)反向传播算法训练的人工神经网络,以壳体厚度与直径比和长度与直径比作为输入参数,预测了 PVC 圆柱壳在轴向压缩下的失效模式。此外,人工神经网络在冷弯型钢领域的应用也越来越广泛, Pala^[23]采用人工神经网络方法估计了在纯压缩和纯弯曲下,冷弯型钢 C 型构件的弹性扭曲屈曲应力,并提出了基于人工神经网络的预测方程。Pala 和 Caglar^[24]在后续研究中使用了这些方程,研究了几何参数对扭曲屈曲应力的影响。在另一项研究中, Guzelbey 等^[25]训练了一个人工神经网络算法,使用实验数据集,旨在预测冷弯型钢腹板的破坏,结果表明,与当前设计规范获得的预测相比,人工神经网络可以提供更准确的预测结果。Degtyarev^[26]训练了人工神经网络模型,以预测冷弯型钢 C 型构件的弹性剪切屈曲荷载和极限抗剪强度,并证明了这比代码规定的设计方程更准确。为了更好地选取适应这种预测的算法,在后续研究中, Degtyarev 和 Naser^[27]比较了 5 种不同的机器学习增强算法在预测冷弯薄壁钢构件的弹性剪切屈曲荷载和剪切强度方面的表现,结果表明, CatBoost 算法在这些增强算法中能够提供最准确的预测。在神经网络模型的研究与开发中,其性能优化的核心在于运用最优化理论指导下的损失函数最小化策略。损失函数,作为衡量模型预测结果与实际观测值偏差的数学工具,常见形式包括均方误差(mean squared error, MSE)和平均绝对误差(mean absolute error, MAE)。最优化理论为神经网络的训练过程提供了数学基础,指导算法调整模型参数,以期达到损失函数的最小值。因此,选择合适的训练算法,不仅是执行具体最优化过程的关键,而且会决定神经网络模型的最终性能。这一过程涉及到复杂的数学计算,包括梯度下降(GD)、拟 Newton 法(BFGS)等方法,旨在找到损失函数的全局最小点或可接受的局部最小点,从而实现模型的最优预测效果。

在本研究中,我们采用人工神经网络对冷弯型钢 C 型构件的弹性临界屈曲荷载进行了预测,所使用的数据集包含了 260 000 根构件的截面参数(输入)及其对应的弹性屈曲荷载(输出),为了有效地训练和验证

神经网络模型,这些数据被划分为训练集、验证集和测试集.在该神经网络的训练过程中,面对的主要挑战之一是解决非线性最小二乘问题,即找到一组参数,使得模型输出与实际输出之间的 MSE 最小.该问题的非线性特性来源于神经网络模型的结构和激活函数,使得优化过程变得复杂和具有挑战性.本文采用了 6 种不同的神经网络训练算法进行模型训练,包括 GD、L-M 算法、带动量的梯度下降(GDM)、可变学习率的梯度下降(GDX)、量化共轭梯度下降(SCG)和 BFGS.这些算法根据其在解决非线性最小二乘问题中的特性和适用性被选中,以适应结构工程领域中的复杂问题.本研究的目的在于全面评估这些算法对人工神经网络性能的影响,并探索它们在预测特定屈曲载荷中的有效性,为结构工程相关的机器学习应用提供参考.

1 神经网络的数据集

1.1 有限条法

有限条法是一种分析薄壁构件屈曲行为的数值方法.在有限条法中,薄壁构件被划分成若干纵向条带,这些条带的纵向边界称为节点线.每个节点线可以通过一个局部坐标系 $oxyz$ 来描述,其中 x 是沿着节点线的方向, y 和 z 是节点线的垂直方向,它们也可以通过一个全局坐标系 $Oxyz$ 来描述.每个节点线有 4 个自由度:在节点线端点的纵向 x 位移 u ,节点线端点的平面内 y 位移 v ,节点线中点的平面外 z 位移 w ,以及绕 x 轴的旋转 θ .这些位移和旋转代表了条带之间的面内膜位移和由板面外弯曲引起的变形.构件的弹性和几何刚度矩阵 K 和 G 是通过组装每个条带的局部弹性和几何刚度矩阵得到的,这些矩阵用于构建稳定特征值问题:

$$(\mathbf{K} - \lambda \mathbf{G}) \cdot \mathbf{v} = \mathbf{0}, \quad (1)$$

其中 \mathbf{v} 表示屈曲的特征向量, λ 表示屈曲的特征值.求解多个给定屈曲半波长 L 的特征值,从而得到弹性屈曲荷载因子和对应半波长的曲线,曲线的最低点通常被认为是构件的临界屈曲点^[6].基于有限条法理论的软件 CUFSM 可以实现对构件进行特征值屈曲分析,得到以屈曲半波长为自变量绘制的临界屈曲荷载因子的曲线图,即为构件的特征曲线,见图 1(a).

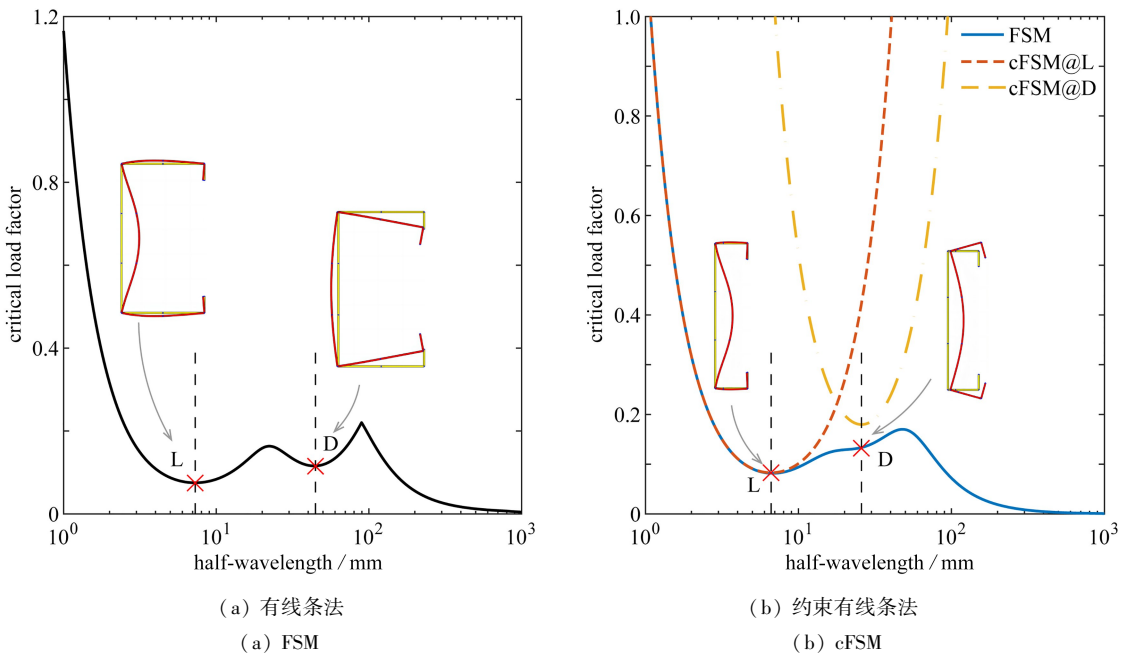


图 1 特征曲线

Fig. 1 Signature curve

注 为了解释图中的颜色,读者可以参考本文的电子网页版本,后同。

1.2 约束有限条法

通常特征曲线有两个最低点,分别代表局部屈曲和畸变屈曲的弹性临界荷载因子和对应的半波长,由于构件截面的多样性,有的构件畸变特征不明显,特征曲线只有一个最低点,因此需要利用约束有限条法进行分析.约束有限条法最初是从半解析有限条法中推导出来的,基于力学准则^[7,28],将有限条一般位移场划分

为4个基本变形类,整体屈曲(G)、畸变屈曲(D)、局部屈曲(L)、剪切和轴向伸缩(ST/O)。约束有限条法可以实现模式分解的功能,可以将一般位移场限制为仅发生期望的屈曲,如图1(b)。其中,黄色虚线代表分解的畸变屈曲的特征曲线,橙色虚线代表分解得到的局部屈曲的特征曲线。由上述所得到的弹性屈曲临界荷载因子结合截面面积 A 和屈服强度 f_y ,可以得到局部和畸变弹性屈曲临界荷载 P_{cr} :

$$P_{cr} = \lambda \times f_y \times A. \quad (2)$$

1.3 数据集

根据冷弯技术指南钢框架产品(SFIA)中对于冷弯型钢C型构件的标准截面库,在截面尺寸限值内随机生成截面尺寸(表1),生成260000个独特的C型钢构件截面数据集。这些数据涵盖了截面卷边 D 、翼缘 B 、腹板 H 和厚度尺寸 t ,由于冷弯型钢构件截面厚度较小,所有的有限条模型均采用中心线尺寸,即 $t/2$ 位置的尺寸(图2),记为腹板 H_c 、翼缘 B_c 、卷边 D_c 、厚度 t 。

表1 冷弯型钢C型截面尺寸范围(单位: mm)

Table 1 Ranges of sectional dimensions of cold-formed lipped channel steels (unit: mm)

component	min	max
H_c	41.15	406.40
B_c	31.75	88.90
D_c	4.78	25.40
t	0.48	3.15

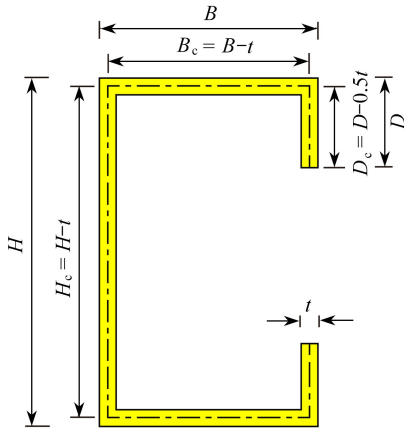


图2 冷弯型钢C型构件截面尺寸示意图

Fig. 2 Schematic diagram of sectional dimensions of cold-formed lipped C-type channel steel columns

对于截面尺寸的参数进行统计分析如图3所示,将4个截面变量制成直方图,并且叠加了核密度估计(KDE)。直方图通过颜色编码直观地表示频率分布,红色轮廓线的核密度估计提供了一个平滑的曲线,以突出中心趋势和分布的扩散。可见,所有变量的频率分布都呈现出平滑且均匀形态,核密度曲线描绘了数据分布的中心趋势和变异。可以看到,260000个随机生成的截面参数样本在整个尺寸空间中均匀分布,对于确保全面分析冷弯型钢构件屈曲的各种可能性至关重要。

构件的边界条件均设置为两端铰接,由CUFSM对所有有限条法模型进行弹性屈曲分析,以确定弹性局部屈曲(P_{crL})和弹性畸变屈曲(P_{crD})的临界荷载及其相应的半波长(λ_{crL} 和 λ_{crD})。此外,我们还计算了无量纲参数,如腹板的板件宽厚比 H_c/t 、翼缘的板件宽厚比 B_c/t 、卷边的板件宽厚比 D_c/t 、截面长宽比 H_c/B_c 和翼缘卷边比 B_c/D_c ,以进一步了解冷弯型钢截面与 P_{crL} 、 P_{crD} 、 λ_{crL} 和 λ_{crD} 有关的几何特性。

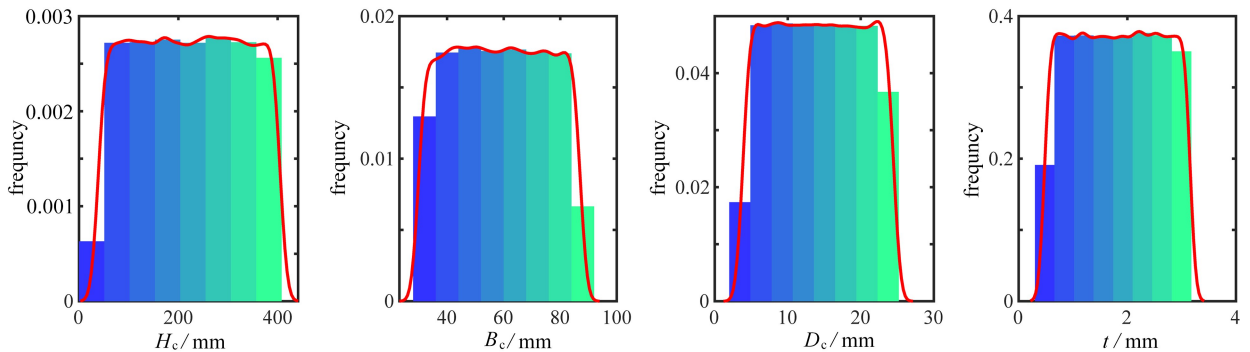


图3 截面尺寸直方图

Fig. 3 Histograms of sectional dimensions

2 反向传播神经网络

本节介绍了神经网络的基本知识,阐述了反向传播神经网络的反向传播算法的基本逻辑,以及 6 种优化算法,并进行网络训练,以便对这 6 种神经网络算法的预测能力进行评估。

2.1 神经网络基础

人工神经网络是机器学习中的一种特殊类型,它是一种前馈多层网络。人工神经网络的主要优势是其训练过程不需要一个预先定义的算法来将输入转化为输出,而是通过一系列代表性的例子进行。人工神经网络的架构受到人脑生物结构的启发^[29-30],由多个相互连接的节点(常称为神经元)组成。这些节点分布在输入层、隐藏层和输出层,每个神经元与下一层中的所有神经元通过权重连接相连。如图 4 所示,人工神经网络通常由 3 种类型的层组成。网络的第一层称为输入层,获取初始输入数据集,输入数据的维度决定了神经网络中输入层的大小。中间层被称为隐藏层,位于网络的输入和输出层之间,隐藏层提高了网络掌握数据中存在的复杂模式和相互关系的能力。网络的最后一层称为输出层,生成网络的输出。

示意图(图 4)展示了一个多层感知器神经网络架构,详细描述了从输入到输出的流程。在人工神经网络的计算过程中,在输入层,输入值为归一化后的值,它们通过权重和偏置连接到隐藏层,然后通过激活函数传递到下一层,同时加上一个偏置值。这一过程持续进行,直到最后一个隐藏层,最终到达输出层。需要注意的是,节点之间的连接不形成闭环,信息在网络中是单向流动的,这就是前向传播算法。具体传播过程如下:

$$a_j^l = f\left(\sum_k w_{ik}^l a_k^{l-1} + b_j^l\right), \quad (3)$$

其中, a_j^l 的值被称为激活值; f 是激活函数,它决定了一个神经元在接受输入后产生的输出值,通常涉及到非线性变换; w_{ik}^l 参数表示第 $l-1$ 层的第 k 个神经元到第 l 层的第 i 个神经元的权重; b_j^l 表示第 l 层的第 j 个神经元的偏置。

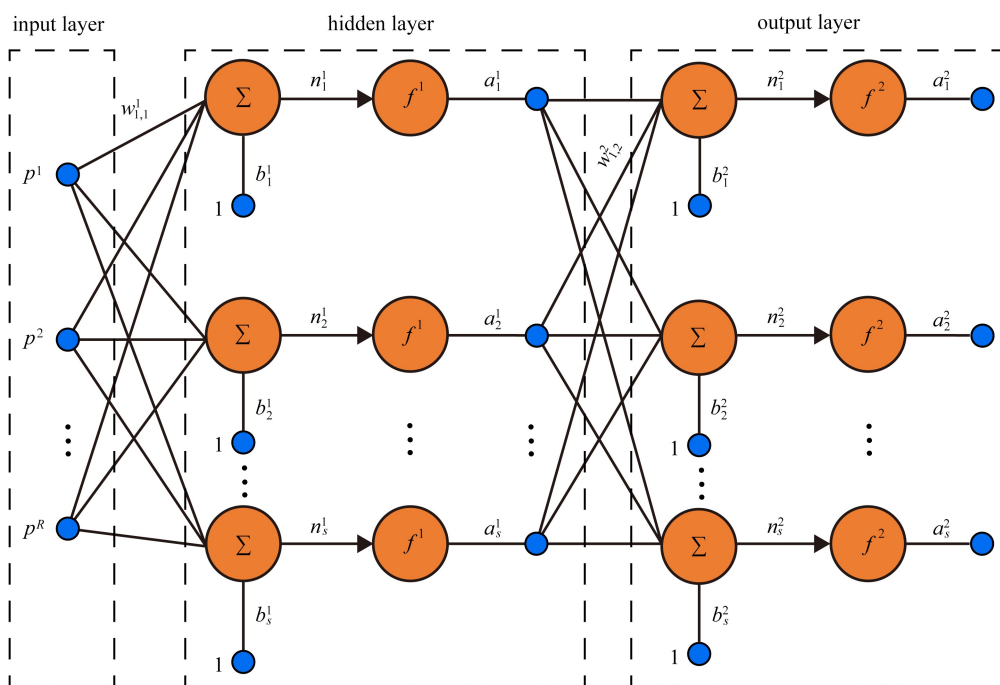


图 4 神经网络架构图

Fig. 4 The neural network architecture

2.2 反向传播

预测模型采用的是反向传播神经网络,通过反向传播算法来训练网络。当输入值被归一化时,输出值也将被归一化,并且需要反归一化,以获得可与目标值进行比较的预测值。预测冷弯型钢构件的弹性屈曲临界荷载本质上是一个回归问题,根据输入变量,如构件的截面特性,预测连续的输出值,即临界荷载。在这种场

景中损失函数选择 MSE, 作为神经网络评估标准和训练过程中的收敛准则, 能够量化模型预测值与实际目标值之间的差异, 衡量模型的整体预测准确性^[31-32]. 为了更好地判断神经网络的性能, 引入平均绝对百分比误差 (mean absolute percentage error, MAPE) 和决定系数 (*R*-squared, R^2) 作为评估标准^[20]. 它们的表达式分别如下:

$$\delta_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (4)$$

$$\delta_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (6)$$

其中, N 是训练样本的总数, y_i 是第 i 个训练样本的实际值, \hat{y}_i 是神经网络对第 i 个训练样本的预测值, \bar{y} 是所有训练样本实际值的平均值.

在训练神经网络时, 其目标是找到一组参数 (如权重和偏置), 使得 MSE 最小化, 解决这个非线性最小二乘问题是最为关键的. 权重和偏置最初在训练过程开始时通常是假定的, 接着通过调整神经网络的权重和偏置值, 使误差最小化. 这可以通过将误差从输出层传递到人工神经网络的输入层来实现^[33], 根据误差对每个权重的影响程度 (梯度) 被反向传播回网络, 在传播过程中对权重和偏置进行调整, 权重的更新是为了减少预测结果与实际值之间的误差, 这通常通过 GD 或其他优化算法来实现, 这个过程被称为多层前馈神经网络的反向传播. 此外, 不同层之间的最终权重和偏置值可以用来量化输入参数对输出参数的影响.

2.3 优化算法

在反向传播神经网络中, 误差反向传播的算法也被称为优化算法, 优化算法的选取决定模型的准确性和收敛速度, 因此基于预测构件的弹性屈曲荷载和半波长, 通过以下 6 种流行的优化算法构建神经网络模型, 对其预测的能力进行评估, 分别为 GD、GDM、GDx、BFGS、L-M 和 SCG, 对比这 6 种优化算法的效率和性能差异, 以选取最佳算法进行更深入的调整.

GD^[34] 是最基本的优化算法, 用于按照损失函数 (误差) 的负梯度方向更新参数, 以减少损失. 在每次迭代中, 参数都会按照梯度和学习率的乘积进行更新:

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} - \alpha \cdot \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_{\text{old}}), \quad (7)$$

其中, $\boldsymbol{\theta}$ 是参数, α 是学习率, L 是损失函数 (误差), $\nabla_{\boldsymbol{\theta}} L$ 是损失函数相对于参数的梯度.

GDM 算法^[34] 在传统的 GD 基础上增加了“动量”. 这种方法考虑了前一步的更新, 允许参数在梯度的方向上积累速度, 从而加速学习过程, 有助于超越局部最小值和平缓区域.

$$\mathbf{v}_t = \beta \mathbf{v}_{t-1} + \alpha \cdot \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}), \quad \boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} - \mathbf{v}_t, \quad (8)$$

其中, \mathbf{v}_t 是当前步的速度, β 是动量系数, 通常设为接近 1 的值.

GDx 算法^[35] 会自动调整学习率, 通常是基于性能的改进或恶化来增加或减少学习率. 这有助于在学习过程中更加灵活地进行参数更新, 避免学习率固定带来的问题.

BFGS^[36] 用于非线性优化问题, BFGS 使用损失函数的梯度来构建损失函数的 Hesse 矩阵的近似, 用于参数更新. 这个方法通常比简单的 GD 更快收敛, 尤其在目标函数比较复杂时.

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} - \alpha_t \cdot \mathbf{H}^{-1} \cdot \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_{\text{old}}), \quad (9)$$

其中, \mathbf{H} 是 Hesse 矩阵的近似, α_t 是在迭代步骤 t 的学习率.

L-M 算法^[34,37] 是一种结合了 GD 和 Gauss-Newton 优化方法的算法, 尤其适用于非线性最小化问题. 当误差减少时, 它更接近于 Gauss-Newton 法; 当误差增加时, 它向 GD 法过渡. 这使得 L-M 算法在神经网络训练中比纯粹的 GD 法更快收敛.

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} - [\mathbf{J}^T \mathbf{J} + \lambda \cdot \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{r}, \quad (10)$$

其中, \mathbf{J} 是损失函数的 Jacobi 矩阵, \mathbf{r} 是残差向量, λ 是调整步长的参数.

SCG^[38]是一种有效的二阶优化算法,不需要显式地计算二阶导数(Hesse 矩阵).它通过使用共轭方向而不是梯度方向来更新参数,可以加速收敛速度,尤其适用于大规模问题。

3 网络训练

数据准备是开发人工神经网络体系结构的第一步,在第 2 节中已经详细阐述了数据集的来源和计算过程.下一步是设计人工神经网络的体系架构,推导出适应于本文预测目标的人工神经网络模型,需要引入一些参数,分别是输入参数、隐藏层内神经元的数量、激活函数和输出参数。

人工神经网络的数据集总数为 260 000 个,其中训练集和验证集百分比之和为 80%,测试集的比例为 20%,基础的网络架构采用由 3 层组成的 9-15-4 配置.第一层的输入参数有 9 个,代表了 C 型钢构件截面的几何特性,包括腹板 H_c 、翼缘 B_c 、截面卷边 D_c 、厚度 t .同时,考虑了板件宽厚比等对弹性屈曲临界值的影响,将腹板的板件宽厚比 H_c/t ,翼缘的板件宽厚比 B_c/t ,卷边的板件宽厚比 D_c/t ,截面长宽比 H_c/B_c ,翼缘卷边比 B_c/D_c 也作为神经网络的输入参数.为了中和数据点之间的差异,需要对数据集进行归一化处理,首先将数据集在 $-1 \sim 1$ 的范围内归一化,然后作为输入提供给神经网络的输入神经元.归一化公式如下:

$$x' = \frac{2(x - \min(x))}{\max(x) - \min(x)} - 1, \quad (11)$$

其中, x 为原始值, x' 为归一化值, $\min(x)$ 和 $\max(x)$ 分别为数据中的最小值和最大值。

第二层隐藏层神经元的数量设置为 15 个,输出层参数为 4 个,分别为局部屈曲弹性临界载荷 P_{crL} , 对应半波长 λ_{crL} , 以及畸变屈曲弹性临界载荷 P_{crD} , 对应半波长 λ_{crD} .选择的网络架构如图 5 所示.使用 MATLAB R2022b 的统计与机器学习工具箱^[39]进行神经网络模型训练。

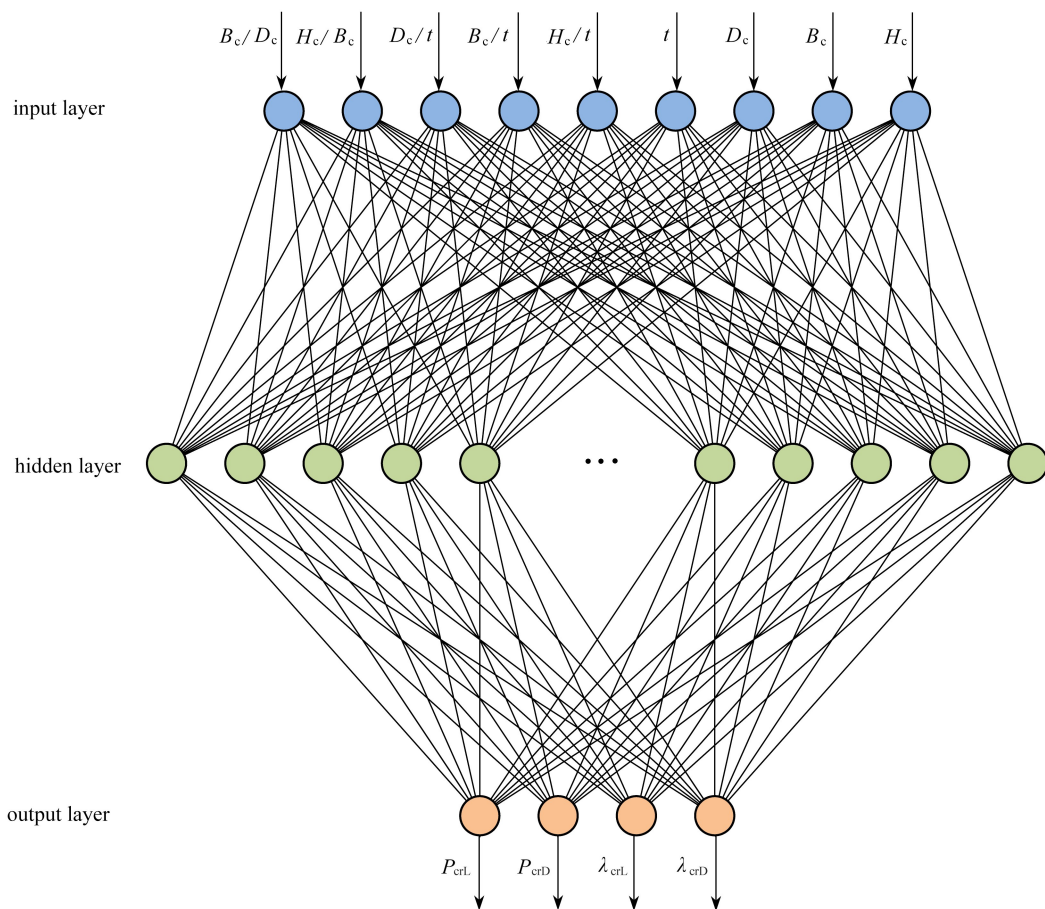


图 5 神经网络(输入层、隐藏层、输出层)

Fig. 5 The neural network (input layer, hidden layer, output layer)

输入层至隐藏层的激活函数采用双曲正切 \tansig 激活函数(式(12)),此函数的数学特性使得其输出能够被有效地限制在 $-1 \sim 1$ 的范围内,这有助于网络在训练过程中保持数值稳定,而隐藏层至输出层使用 purelin 线性激活函数(式(13)),使网络直接输出预测值,而不对其进行任何非线性变换:

$$\tan \operatorname{sig}(x) = \frac{2}{1 + e^{-2x}} - 1, \quad (12)$$

其中, x 是输入值, e 是自然常数.

$$y = x, \quad (13)$$

其中, y 是激活函数的输出, x 是激活函数的输入.

采用反向传播算法对人工神经网络进行训练,在训练过程中,向网络提供一组训练样本,并使用 GD 或其他优化算法迭代调整权重和偏差.训练过程旨在最小化实际数据集与人工神经网络预测结果之间的误差,提高网络做出准确预测的能力.在计算过程中,归一化后的输入向量生成一组相应的输出值,随后,给定输出和预测输出之间的误差,通过网络向后传播.在整个过程中, MSE 被减小,最终经过训练后的人工神经网络的预测输出值接近目标输出值.

4 训练算法的性能

不同的优化算法基于各自独特的原理,应用于不同问题时,它们的性能可能会有显著差异.某一算法在特定问题上可能展现出超越其他算法的预测准确性,而在另一问题上则可能性能不佳.因此,在特定问题上寻找最合适的算法及其最优超参数组合是至关重要的.这一过程不仅涉及算法选择,还包括精细调整超参数,以确保所选算法在特定应用场景下能够发挥最佳效能.针对这 6 种训练算法研究其性能和效率,采用一致的网络架构(9-15-4)和激活函数(\tanh 函数),学习率设置为 0.01,网络的初始权重和偏置由 Gauss 分布随机生成,分别使用 GD、GDX、GDM、L-M、SCG 和 BFG 优化算法训练网络.同时,为了减小过拟合控制泛化误差,采用早停法,即当验证集误差连续 6 次增大时停止训练.对网络持续训练,直到达到最大 500 次 epoch,使用全部样本完成一次正向传播和反向传播称为一次 epoch.每种训练算法重复训练 20 次,分别统计运行时长和 MSE.根据模型训练结果选取准确性最高的优化算法,进行精细化调整,利用随机网络搜索最终确定超参数.

4.1 训练结果

在表 2 中列出了 6 种算法(GDX、GDM、GD、SCG、BFG 和 L-M)的运行时长,并进行比较,时长越小,表明在对应的性能指标上算法的效率越高.L-M 在所有列出的性能指标上显示了最高的数值,这可能意味着它在计算效率方面效果欠佳.SCG 和 BFG 在表 2 中的性能数值相比 L-M 更低,这表明它们可能在计算效率的性能指标上更优.值得注意的是,GDX 虽然训练时长最短,但是在 500 次 epoch 内出现了过拟合现象,过拟合的特点是模型能够对训练中使用的样本做出良好的预测,而对模型之前未见过的新数据样本做出较差的预测.

表 3 为计算 6 组算法的验证集误差,使用平均 MSE 作为预测准确性的度量标准,数值越小意味着预测误差越小.在这两个表格中,L-M 算法在平均 MSE、最小 MSE 和最大 MSE 上表现出相对较好的性能,这意味着它在数据集上的预测稳定性是这些算法中最好的.在表 3 中,SCG 和 BFG 的 MSE 表现介于 L-M 和其他方法之间,显示出中等的预测稳定性.总的来说,L-M、SCG、BFG 的计算时长最长,验证集误差最小.

表 2 各训练算法运行时长

Table 2 Runtime lengths of various training algorithms

training algorithm	average time/s	ratio	min time/s	max time/s
GDX	1.29	1.00	1.16	1.41
GDM	3.93	3.05	0.32	4.18
GD	4.13	3.21	4.10	4.19
SCG	9.36	7.27	7.81	37.75
BFG	24.80	19.26	16.93	167.68
L-M	132.10	102.62	110.54	135.84

表 3 各训练算法验证集误差

Table 3 Validation set errors for various training algorithms

training algorithm	average MSE	ratio	min MSE	max MSE
L-M	2.01E4	1.00	1.67E4	2.56E4
BFG	4.48E4	2.16	3.89E4	5.74E4
SCG	5.69E4	2.74	4.95E4	6.74E4
GDX	8.73E3	42.10	6.20E3	1.30E2
GD	8.67E2	417.85	6.92E2	1.16E1
GDM	8.99E2	433.52	5.09E2	2.02E1

根据表 2、3 选择 L-M、SCG 和 BFG 3 种训练算法,将 SCG、BFG 和 L-M 最大 epoch 改为 5 000,其他网络参数保持不变,重新训练 20 次.在进行算法效率比较时(表 4),可以得到 L-M 算法在训练时需要的平均时间远多于 BFG 和 SCG.虽然这增加了训练成本,但从表 5 中可以得知,这种投入与 L-M 在验证集上展现出的高精度是一致的,平均 MSE 最低,仅为其他两种算法的 0.57 倍.

值得注意的是,即使将 SCG 和 BFG 的最大 epoch 设为 5 000,由于早停法的设置,这两种训练算法都不能跑满整个 epoch,说明 SCG、BFG 都出现了过拟合的现象.L-M 算法能够运行到设定的最大 epoch(见表 6),没有出现显著的过拟合现象,能够实现更优的性能,这说明在长时间的训练中,L-M 算法能够进一步优化其预测准确度.因此最终选取最优训练算法 L-M,对神经网络进行精细化调整.

表 4 BFG 和 SCG 训练算法时长对比

Table 4 Comparison of training durations between BFG and SCG algorithms

training algorithm	average time/s	ratio	min time/s	max time/s
BFG	40.23	1.00	22.46	64.90
SCG	45.27	1.13	19.40	208.06
L-M	363.35	9.03	107.76	1 359.262

表 5 BFG 和 SCG 验证集 MSE 对比

Table 5 Comparison of validation set MSE between BFG and SCG algorithms

training algorithm	average MSE	ratio	min MSE	max MSE
BFG	3.40E4	1.00	2.40E4	8.6E4
SCG	3.40E4	1.00	2.70E4	4.5E4
L-M	1.95E4	0.57	1.68E4	2.33E4

表 6 BFG 和 SCG epoch 对比

Table 6 Comparison of epochs between BFG and SCG algorithms

training algorithm	average epoch	ratio	min epoch	max epoch
BFG	1 191.90	1.00	657	1 957
SCG	2 358.55	1.98	1 247	4 185
L-M	1 344.75	1.13	394	5 000

4.2 L-M 算法

在神经网络训练中,随机网格搜索是一种超参数优化方法,旨在通过随机采样超参数空间来找到最佳的模型超参数配置.超参数是在学习过程开始之前设置的参数,不同于模型训练过程中学习的参数(例如,权重和偏置).超参数直接控制着模型的行为和性能(例如,学习率、隐藏层的数量、每层的神经元数等).相对于穷尽性的网格搜索,当搜索空间庞大时,此过程提供了一个切实可行的替代方案.通过随机采样空间,它为给定模型提供了一种成本效益高的方式来逼近最优超参数.

在上述方法中,定义了神经网络的超参数空间,隐藏层的神经元数量从 2 到 128,以 2 为增量变化,学习率设定为 0.000 1,0.001,0.01 和 0.1.采用随机网格搜索来选择神经元数量和学习率的组合,确保至少覆盖总超参数空间的 25%.为了防止过拟合,如果验证集上的 MSE 连续 6 次迭代增加,则停止训练.一个 epoch 包括所有样本通过网络的完整前向和后向传播,最大 epoch 数设为 1 000.激活函数在隐藏层使用 tanh 函数,而隐藏至输出层使用 purelin 函数.最优超参数是那些使验证集性能达到最佳的参数,这些参数被指定为模型的最优超参数.

图 6 给出了通过随机网格搜索得出的不同超参数组合对神经网络训练效果的影响,具体表现在运行时

间和 MSE 两个性能指标上.可以看出,随着神经网络中神经元数量的增加,模型的 MSE 呈现出逐渐下降的趋势,表明了预测准确性的提高.与此同时,模型的计算时间亦随之增长,这反映了提升模型性能所需的额外计算代价.经过多次迭代的训练和验证后,最小的 MSE 为 8.56×10^{-5} ,对应的最优超参数是学习率为 0.001,隐藏层神经元个数为 124,训练时长为 21 572.6 s.这组超参数在验证集上展现出了最佳的性能,因此被认为是模型的最优超参数.

考虑到随机初始权重和偏置在神经网络训练中所带来的局限性,通常的做法是使用最优超参数设置多次训练网络.这种做法有助于减少因随机初始条件而引起的性能波动,从而提供更为可靠的网络性能评估.通过使用相同的最优超参数重复训练过程,可以选择表现最佳的模型结果,这提高了神经网络模型的预测稳定性和泛化能力.

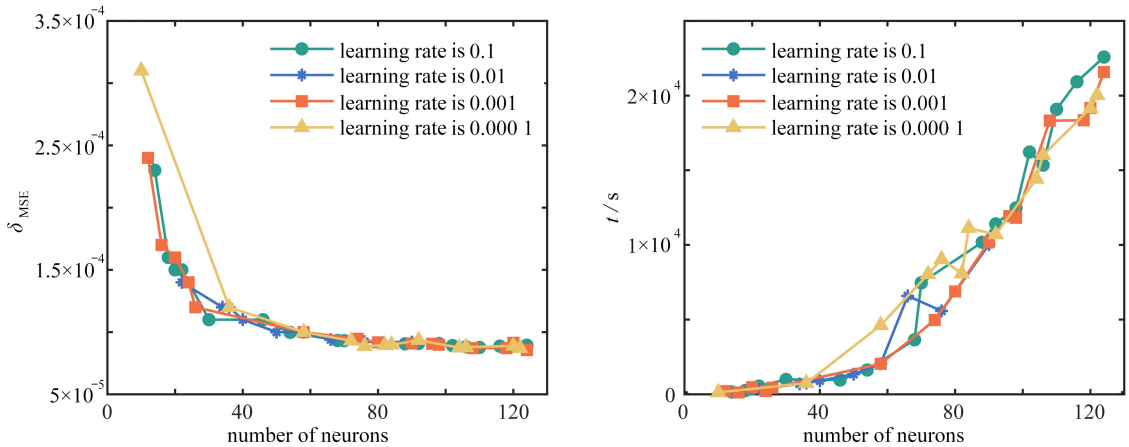


图6 MSE 和计算时长随神经元个数的变化

Fig. 6 MSE and computation time changes with the number of neurons

4.3 网络评估

图7将反向传播神经网络的预测结果与使用CUFSM方法计算出的不同屈曲载荷的数值联系起来,通过线性回归分析绘制拟合线(fit line).图7(a)和7(b)分别显示了临界局部(P_{crL})和扭曲(P_{crD})屈曲载荷的关系,而图7(c)和7(d)则对应于局部(λ_{crL})和扭曲(λ_{crD})屈曲的半波长.图中包括回归拟合线(fit line)、完美预测线(1:1线),并显示回归方程、决定系数(R^2)、相关系数(R)和MAPE.颜色梯度表示拟合线周围数据点的密度,密度越高的区域表示数据点越集中.可以看出,拟合线与1:1线几乎重合,这意味着每一个预测值都与相应的实际值完全匹配,接近1的 R 值和 R^2 值以及较低的MSE百分比表明,反向传播神经网络模型具有出色的预测性能和高度的准确性.观察结果表明,以MAPE衡量的模型预测性能在估计 P_{crL} 和 P_{crD} 时显示出较大的差异,与半波长参数 λ_{crL} 和 λ_{crD} 相比,MAPE值分别为5.3436%和1.5533%.这可能表明,在预测 P_{crL} 和 P_{crD} 现象方面,模型的敏感性或复杂性存在差异.

4.4 GPU加速训练神经网络

L-M算法的核心为迭代式的非线性最小化过程,其顺序性质和对前一迭代结果的依赖显著,限制了并行化潜力.该算法在每次迭代中需计算Jacobi矩阵、Hesse矩阵的近似及其逆矩阵,这些稀疏矩阵的处理在GPU上并非最优化,特别是考虑到其尺寸和稀疏性的变化性.此外,算法中的动态参数调整和复杂条件判断进一步加剧了其并行化的困难,使得L-M算法无法充分利用GPU的并行处理优势.L-M算法结合了GD法和Gauss-Newton法的优点,在接近解时表现出高度的敏感性和精确度,这使得它在某些特定问题上比基于梯度的方法更准确.L-M算法通常能提供更好的性能,因为它能更精确地调整参数以适应数据.相比之下,一些可通过GPU加速的算法(如SCG),尽管在处理大规模数据集时效率更高,但这并不总是意味着预测结果更准确.

图8为两种不同优化算法在预测准确性方面的性能对比,即支持GPU加速的SCG算法和不支持GPU加速的L-M算法.预测准确性通过MAPE来衡量,较低的MAPE值表明更高的预测精度.从图中可以看出,在预测 P_{crL} 、 P_{crD} 、 λ_{crL} 和 λ_{crD} 4个参数时,L-M算法的MAPE值均低于SCG算法,这表明在预测冷弯型钢构件的

弹性屈曲临界荷载和半波长中,不支持 GPU 加速的 L-M 算法提供了更准确的预测结果,尤其在 P_{crL} 的预测中,L-M 算法的优势更为显著。

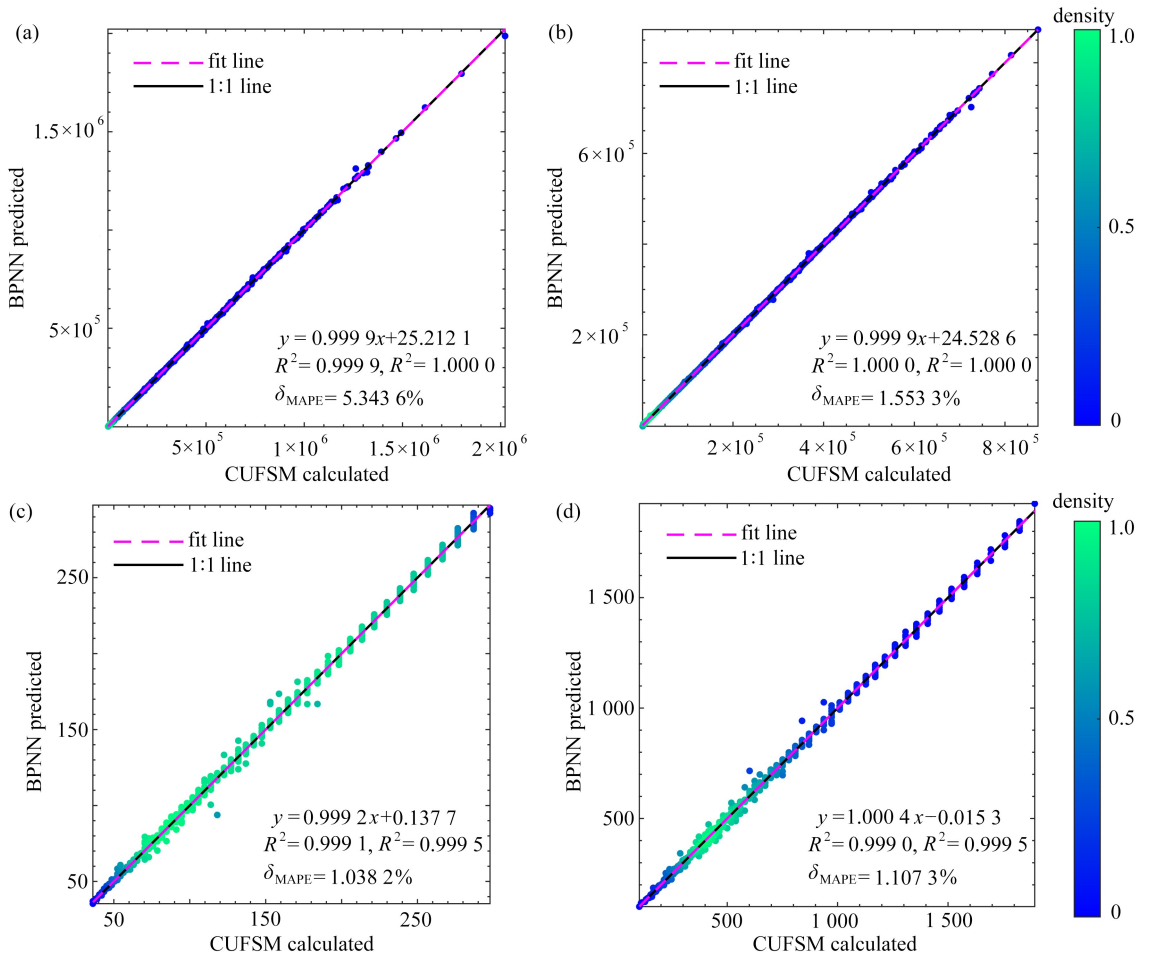


图 7 神经网络模型预测结果与 CUFSM 计算结果对比

Fig. 7 Comparison between neural network model predictions and CUFSM calculation results

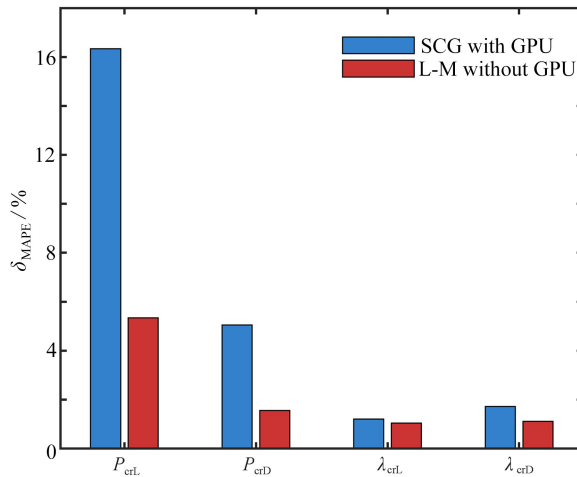


图 8 GPU 加速的 SCG 算法与无 GPU 加速的 L-M 算法预测值的 MAPE 对比

Fig. 8 Comparison of MAPE values between the GPU-accelerated SCG algorithm and the non-GPU-accelerated L-M algorithm

在训练效率方面,L-M 算法完成网络模型训练需要 21 572.60 s,而 SCG 算法需要 5 272.44 s.虽然 GPU 加速算法在处理大规模数据集和实现快速运算方面有明显优势,但 L-M 算法在预测冷弯型钢构件的弹性屈

曲临界荷载和半波长中表现出色,可以提供更高的准确度。

5 结 论

本文构建了一个包含 260 000 个样本的冷弯型钢 C 型截面弹性屈曲临界值数据集,并建立了一个多输出反向传播神经网络模型用于预测弹性屈曲临界值。考虑到这一预测任务属于非线性最小二乘回归问题,搭建了 235 组反向传播神经网络模型,深入探究了 GD、GDX、GDM、BFG、SCG 和 L-M 等不同训练算法对模型性能的影响。综合考量后发现,L-M 算法在非线性最小二乘回归问题上展现出较高的准确性。然而,由于 L-M 算法在迭代过程中依赖 Jacobi 矩阵及其逆矩阵,因此无法利用 GPU 加速。相比之下,其他算法尽管能够实现 GPU 加速,但在准确度方面不及 L-M 算法。未来,我们将继续研究将继续探索既能利用 GPU 加速又能保持高准确度的训练算法。

参考文献(References):

- [1] VON KÁRMÁN T, SECHLER E E, DONNELL L H. The strength of thin plates in compression[J]. *Journal of Fluids Engineering*, 1932, **54**(2): 53-56.
- [2] 杨梦欢. 冷弯薄壁卷边槽钢受压柱畸变屈曲性能的研究[D]. 武汉: 武汉理工大学, 2013. (YANG Menghuan. Distortional buckling behavior of cold-formed thin-wall steel lipped channel columns under compression[D]. Wuhan: Wuhan University of Technology, 2013. (in Chinese))
- [3] 高段. 冷弯薄壁卷边槽钢受弯构件畸变屈曲性能研究[D]. 武汉: 武汉理工大学, 2013. (GAO Duan. Study on distortional buckling of cold-formed thin-walled lipped channel steel in bending[D]. Wuhan: Wuhan University of Technology, 2013. (in Chinese))
- [4] 陈骥. 冷弯薄壁型钢构件的直接强度设计法[J]. 建筑钢结构进展, 2003, **5**(4): 5-13. (CHEN Ji. Direct strength method for the design of cold-formed lipped channel members[J]. *Progress in Steel Building Structures*, 2003, **5**(4): 5-13. (in Chinese))
- [5] CHEUNG Y, THAM L. *Finite Strip Method*[M]. CRC Press, 1997.
- [6] LI Z, SCHAFER B W. Buckling analysis of cold-formed steel members with general boundary conditions using CUFMSM conventional and constrained finite strip methods[C]//*20th International Specialty Conference on Cold-Formed Steel Structures*. St Louis, Missouri, 2010.
- [7] ÁDÁNY S, SCHAFER B W. Buckling mode decomposition of single-branched open cross-section members via finite strip method: derivation[J]. *Thin-Walled Structures*, 2006, **44**(5): 563-584.
- [8] UNGUREANU V, DUBINA D. Recent research advances on ECBL approach, part I: plastic-elastic interactive buckling of cold-formed steel sections[J]. *Thin-Walled Structures*, 2004, **42**(2): 177-194.
- [9] DAVIES J M. Recent research advances in cold-formed steel structures[J]. *Journal of Constructional Steel Research*, 2000, **55**(1/2/3): 267-288.
- [10] WANG T, ZHA Z, PAN C. Prediction for elastic local buckling stress and ultimate strength of H-section beam [J]. *Heliyon*, 2023, **9**(4): e14700.
- [11] ZARRINGOL M, THAI H T, THAI S, et al. Application of ANN to the design of CFST columns[J]. *Structures*, 2020, **28**: 2203-2220.
- [12] WU T Y, EL-TAWIL S, MCCORMICK J. Effect of cyclic flange local buckling on the capacity of steel members [J]. *Engineering Structures*, 2019, **200**: 109705.
- [13] PITTON S F, RICCI S, BISAGNI C. Buckling optimization of variable stiffness cylindrical shells through artificial intelligence techniques[J]. *Composite Structures*, 2019, **230**: 111513.
- [14] TOHIDI S, SHARIFI Y. Neural networks for inelastic distortional buckling capacity assessment of steel I-beams[J]. *Thin-Walled Structures*, 2015, **94**: 359-371.
- [15] 张广江, 杨德泽, 楚锡华. 基于人工神经网络的颗粒材料本构关系及边值问题研究[J]. 应用数学和力学, 2024, **45**(2): 155-166. (ZHANG Guangjiang, YANG Deze, CHU Xihua. Study on constitutive relations and boundary value problems of granular materials based on artificial neural networks[J]. *Applied Mathematics and Mechanics*, 2024, **45**(2): 155-166. (in Chinese))
- [16] KAVEH A, BAKHSHPOORI T, HAMZE-ZIABARI S M. GMDH-based prediction of shear strength of FRP-RC

- beams with and without stirrups[J]. *Computers and Concrete*, 2018, **22**(2): 197-207.
- [17] KAVEH A, ELMIEH R, SERVATI H. Prediction of moment-rotation characteristic for semi-rigid connections using BP neural networks[C]//*Computational Engineering Using Metaphors From Nature*, 2001: 15-24.
- [18] IRANMANESH A, KAVEH A. Structural optimization by gradient-based neural networks[J]. *International Journal for Numerical Methods in Engineering*, 1999, **46**(2): 297-311.
- [19] KAVEH A, IRANMANESH A. Comparative study of back propagation and improved counter propagation neural nets in structural analysis and optimization[J]. *International Journal of Space Structures*, 1998, **13**(4): 177-185.
- [20] MOJTABAEI S M, BECQUE J, HAJIRASOULIHA I, et al. Predicting the buckling behaviour of thin-walled structural elements using machine learning methods[J]. *Thin-Walled Structures*, 2023, **184**: 110518.
- [21] WASZCZYSZYN Z, BARTCZAK M. Neural prediction of buckling loads of cylindrical shells with geometrical imperfections[J]. *International Journal of Non-Linear Mechanics*, 2002, **37**(4/5): 763-775.
- [22] MARKOPOULOS A P, MANOLAKOS D E, VAXEVANIDIS N M. Prediction of the collapse modes of PVC cylindrical shells under compressive axial loads using artificial neural networks[C]//*The International Federation for Information Processing*. Boston, MA, 2007: 251-258.
- [23] PALA M. A new formulation for distortional buckling stress in cold-formed steel members[J]. *Journal of Constructional Steel Research*, 2006, **62**(7): 716-722.
- [24] PALA M, CAGLAR N. A parametric study for distortional buckling stress on cold-formed steel using a neural network[J]. *Journal of Constructional Steel Research*, 2007, **63**(5): 686-691.
- [25] GUZELBEY I H, CEVIK A, ERKLIĞ A. Prediction of web crippling strength of cold-formed steel sheetings using neural networks[J]. *Journal of Constructional Steel Research*, 2006, **62**(10): 962-973.
- [26] DEGTYAREV V V. Neural networks for predicting shear strength of CFS channels with slotted webs[J]. *Journal of Constructional Steel Research*, 2021, **177**: 106443.
- [27] DEGTYAREV V V, NASER M Z. Boosting machines for predicting shear strength of CFS channels with staggered web perforations[J]. *Structures*, 2021, **34**: 3391-3403.
- [28] ÁDÁNY S, SCHAFFER B W. A full modal decomposition of thin-walled, single-branched open cross-section members via the constrained finite strip method[J]. *Journal of Constructional Steel Research*, 2008, **64**(1): 12-29.
- [29] NOORZAEI J, HAKIM S, JAAFAR M, et al. An optimal architecture of artificial neural network for predicting compressive strength of concrete[J]. *Indian Concrete Journal*, 2007, **81**(8): 17-24.
- [30] FARZINPOUR A, DEHCHEHMEH E M, BROUJERDIAN V, et al. Efficient boosting-based algorithms for shear strength prediction of squat RC walls[J]. *Case Studies in Construction Materials*, 2023, **18**: e01928.
- [31] BAI C, NGUYEN H, ASTERIS P G, et al. A refreshing view of soft computing models for predicting the deflection of reinforced concrete beams[J]. *Applied Soft Computing*, 2020, **97**: 106831.
- [32] HAKIM S J S, PAKNAHAD M, KAMARUDIN A F, et al. Buckling prediction in steel columns: unveiling insights with artificial neural networks[J]. *International Journal of Engineering Trends and Technology*, 2023, **71**(9): 322-330.
- [33] AHMAD A, COTSOVOS D M, LAGAROS N D. Assessing the reliability of RC code predictions through the use of artificial neural networks[C]//*School of Energy, Geoscience, Infrastructure and Society Institute for Infrastructure & Environment*. 2016.
- [34] HAGAN M T, DEMUTH H B, BEALE M. *Neural Network Design*[M]. PWS Publishing Co., 1997.
- [35] *The MathWorks*[M]. Natick, MA: MATLAB, 2012.
- [36] DENNIS JR J E, SCHNABEL R B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*[M]. SIAM, 1996.
- [37] MARQUARDT D W. An algorithm for least-squares estimation of nonlinear parameters[J]. *Journal of the Society for Industrial and Applied Mathematics*, 1963, **11**(2): 431-441.
- [38] MØLLER M F. A scaled conjugate gradient algorithm for fast supervised learning[J]. *Neural Networks*, 1993, **6**(4): 525-533.
- [39] WORKS M. *Statistics and Machine Learning Toolbox User's Guide*[M]. Matwork Inc, 2017.