

基于小样本机器学习构建镍基高温合金应力应变本构模型

尹 茸¹, 孙嘉言², 许庆彦²

(1. 小米人工智能实验室, 北京 100102; 2. 清华大学 材料学院先进成形制造教育部重点实验室, 北京 100083)

摘要: 在镍基高温合金力学本构模型构建的过程中, 使用小样本机器学习方法, 结合数据增强、网络结构优化、迁移学习等方法, 构建了小样本神经网络模型, 降低了对实验数据量的依赖性, 经过测试, 模型精度高于一般BP神经网络和唯象型本构模型。

关键词: 高温合金; 应力-应变本构模型; 小样本; 机器学习

高温合金具有较高的高温强度, 良好的抗氧化和抗腐蚀性能, 良好的疲劳性能、断裂韧性等综合性能, 广泛应用于航空、航天、石油、化工、舰船等领域^[1]。在铸件设计阶段, 精准的数值模拟可以极大地提高生产效率, 提高铸件质量。其中, 根据应力场数值模拟的结果, 我们可以调控凝固过程^[2]、预测铸件变形^[3]、热裂缺陷^[4]、力学性能^[5]等。图1展示的是商用软件对高温合金涡轮叶片应力场的模拟结果, 通过不断调整铸件设计及凝固工艺参数、降低应力集中情况, 尽可能避免相关缺陷的发生。

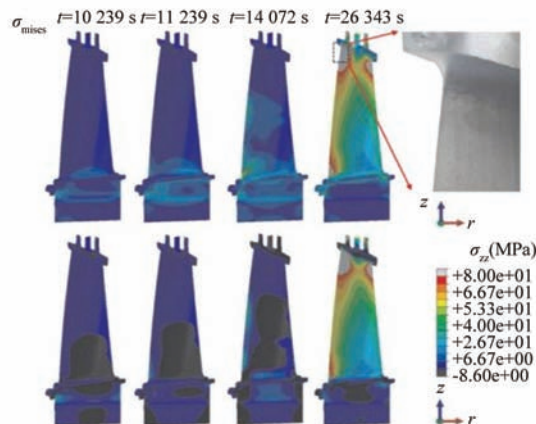


图1 高温合金涡轮叶片应力场的数值模拟结果

Fig. 1 Numerical simulation results of stress field in superalloy turbine blades

截至目前, 常见铸造商用软件尚无法实现针对实际铸件的应力应变场分布的准确数值模拟, 且难以进一步优化, 而常用有限元力学数值模拟软件缺乏针对所需材料的准确力学性能模型, 对数值模拟结果准确性造成一定影响。为实现凝固过程应力场的准确预测, 需要针对材料的高温力学行为开展研究, 建立准确的力学本构模型。本构模型的建立是极具挑战的, 需要根据生产中遇到的问题, 不断增加输入参数, 并对尚未发生的过程有优秀的泛化预测能力。

建立应力-应变本构模型的经典方法是Arrhenius模型, 这是Sellars和Tegart建立的一种本构模型广泛用于描述材料热变形过程中变形条件和流变应力之间的关系^[6], 一般用相对固定的拟合过程, 求解应力与温度、应变速率、应变量的表达式。

作者简介:

尹茸 (1988-), 男, 硕士, 研究方向为人工智能、智能制造。E-mail: yinrong@xiaomi.com

通讯作者:

许庆彦, 男, 博士, 教授, 博士生导师。电话: 13701087143, E-mail: scjxqy@mail.tsinghua.edu.cn

中图分类号: TG146.2

文献标识码: A

文章编号: 1001-4977(2023)

09-1091-08

基金项目:

国家科技重大专项 (2017-VII-0008-0101)。

收稿日期:

2022-09-12 收到初稿,

2023-06-27 收到修订稿。

$$\sigma = \frac{1}{\alpha(\epsilon)} \ln \left[\left(\frac{Z}{A(\epsilon)} \right)^{-\frac{1}{n(\epsilon)}} \left(\left(\frac{Z}{A(\epsilon)} \right)^{\frac{2}{n(\epsilon)}} + 1 \right)^{\frac{1}{2}} \right] \quad (1)$$

$$Z = \epsilon \exp[Q(\epsilon)/(RT)] \quad (2)$$

式中： n 、 A 、 Q 、 α 通过多项式拟合得到， Z 为Zener-Hollomon参数。

Arrhenius模型的缺点是，没有考虑材料在不同方向、温度等变形条件下的变形机制，需要分场景建模，工作量巨大；另外，拟合过程不允许关键数据点的缺失，实验成本大幅增加。

BP神经网络（见图2）也是常用的方法。神经网络方法具有独特的优势，可以灵活地增加多种输入参数，完全不受公式的约束。BP神经网络由输入、输出及隐含层构成。输入层对应实验中的自变量，输出层对应因变量，隐含层包含大量节点，通过神经元与激活函数等结构，将输入映射为输出。训练时数据向前传递，计算误差，对误差求导得到每个网络参数的梯度，根据梯度调节参数。重复这个过程，调整网络参数直到误差最小，就完成了网络的训练。BP神经网络的缺点是必须用大量数据训练才能达到良好的精度，而铸造领域往往只能获取到分散的小数据集。

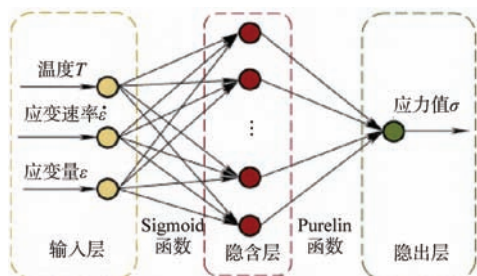


图2 BP神经网络
Fig. 2 BP neural network

实验数据不足是材料力学模型建立过程中经常面临的问题。高温合金铸件的生产成本高、生产过程复杂，受实际因素限制，模型试验数据往往较少，导致建立相关计算模型变得困难。小样本通常指样本不足以约束潜在模型的情况，过少的样本数会导致分类线不受约束，从而得到错误的分类结果；而增加采样点

后，分类线获得更多约束，这将提高模型的准确率^[7]。图3为函数 $y = \sin(x)$ 在不同样本约束模型下的拟合结果图，由图中可以清晰地看出，随着样本数的增加，拟合结果不断优化，在样本数量达到一定量时，拟合结果已基本接近理想模型状态。

在构建DZ411应力-应变模型的过程中，就存在样本数量不足（即小样本）问题。在建模过程中，将单一实验条件下获得的一组力学响应曲线定义为单个样本，样本数量不足100个，符合小样本情况。我们的目标是得到精度尽量高的模型，但在数据极少的情况下，很难得到理想模型。在这种小样本场景下，无论Arrhenius方法还是BP神经网络方法，模型精度都无法满足应力场精准模拟的要求。

解决小样本学习的主要方法是迁移学习。现实中，这些小样本背后蕴含着人类理解的规律，这种规律包括公式、定理、模式等。小样本学习的核心问题是如何将小样本与这些规律联合从而得到精度更高的模型。以图3为例，最理想的情况下，如果拟合模型的程序“知道”正在拟合一个sin函数，那么只需要极少样本就能确定sin函数的周期、相位，完成目标模型的拟合。

在构建DZ411应力-应变模型的过程中，使用小样本机器学习的方法构建了应力-应变本构模型，该方法需要综合运用特定的数据获取方法、神经网络结构、迁移学习方法。

1 原理与方法

1.1 数据获取

1.1.1 数据分布

输入特征符合正态分布可以提升模型精度^[8]。设计实验参数，或者筛选数据的时候，我们统计每个区间 $[x, x+t]$ 的样本个数 N_i 。 $N_i = \text{ceil}(Me^{-\frac{x^2}{2}})$ ，其中 M 为大于0的整数， t 是样本空间均分后每个区域的大小。另一方面，如果我们获得的实验数据无法满足这个分布，就可以尝试将其映射为正态分布。我们可以使用指数变换把旧的特征值 x 变换为 x_{new} 。

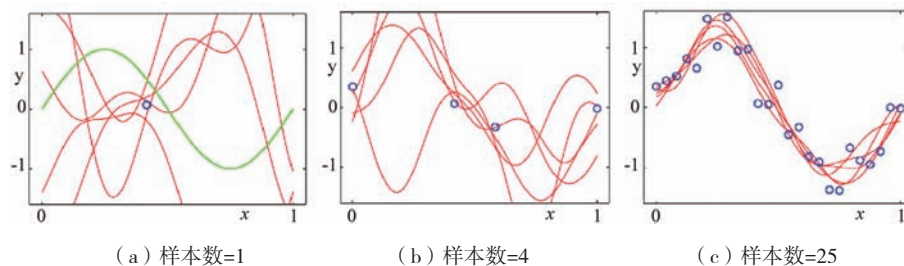


图3 不同样本数量下 $y = \sin(x)$ 拟合结果

Fig. 3 Fitting results of $y = \sin(x)$ for different number of samples

$$x_{\text{new}} = \begin{cases} \frac{(x+\gamma)^\beta}{\|(x+\gamma)^\beta\|_2} & \text{if } \beta \neq 0 \\ \frac{\log(x+\gamma)^\beta}{\|\log(x+\gamma)^\beta\|_2} & \text{if } \beta = 0 \end{cases} \quad (3)$$

式中： $\gamma = \max(x_i)$ ；保证 $x+\gamma$ 为正数； β 是超参数，一般不等于1，可以设为0.5作为初始值； x 是特征值。

应变速率、真应力、变形温度等值数量级相差较大，在符合正态分布之后，还应该使用归一化处理，将特征值映射到[0, 1]区间内，从而提高模型精度。可以尝试线性归一化、均值归一化、排名归一化等处理方法，这些方法可能适配不同的数据情况，在处理速度、模型精度会有所不同。

使用贝叶斯方法^[9]等主动优化的方法，根据每一次实验结果，再计算新的实验参数，比正态分布的效果更好，但是很多场景下，分批次的实验可能是不现实的，这样就不能采用这种方法。

1.1.2 异常数据

通常，要去除异常数据。异常数据对拟合结果的影响很大，将异常点去除后，拟合效果明显增强。常见的方法包括可靠协方差^[10]、支持向量机分类、随机森林分类、局部离群系数^[11]。但在小样本场景下，每个数据都至关重要，不能粗暴地去除一个数据点。我们应该结合专业经验，分析这个数据离群的原因。如果没有足够的理由将数据归类为错误，我们可以将数据保留、不保留分别建模，然后通过测试结果再决定离群数据是否保留。

此外，特征选择也是常用的方法，即从数据中去除没有贡献的特征。常见的方法包括目标相关性分析、重要度分析、递归去除。但在小样本、特征列比较少的情形下，剔除一列特征将缺失大量信息。可以将一行数据中没有贡献的特征值标记为缺失，一般是更好的策略。

1.1.3 数据增强

小样本场景下，数据增强是有效的提升神经网络性能的方法。以SMOTE数据增强方法为例，对每一个特征 x ，按欧氏距离找出 K 个最近邻样本，从中随机选取一个样本点，假设选择的近邻点为 x ，可以计算新增的点为：

$$x_{\text{new}} = x + \text{rand}(0, 1) \times (x_{\text{new}} - x) \quad (4)$$

用更复杂的曲线代替直线型数据增强、添加噪声、使用对抗神经网络、自动数据增强都是有用的方法。

除了数据行的增加，数据列也是可以增加的。模型训练中，经常增加的特征包括：与输入关联的材料

学知识、输入特征值的公式变换项。

1.1.4 测试集选取

测试集是评判模型精度的数据集，正确的选取才能保证模型训练的有效性。理想情况下，测试样本也是从样本真实分布中独立同分布采样得到，测试集应该尽可能与训练集互斥。

在小样本场景下，简便的方法是正态分布空间上概率最大的区域，选取覆盖全部输入参数的采样点作为测试集，并确保不参与训练。另外，如果计算资源充足，也可以使用交叉验证法，具体的，将一个 (ϵ, T) 条件下的实验数据作为一个数据集，得到 k 个数据集，每次将其中一个数据集作为测试集，剩下 $k-1$ 个数据集作为训练集进行训练。重复 k 次训练和评估，取平均值得到模型精度。

$$E_j = \max \frac{1}{k} \sum_{i=1}^k E(\epsilon_i, T_i, \sigma_i, H_j) \quad (5)$$

式中： H_j 是模型的超参数集合； E 是最优模型的测试结果； E_j 是模型训练中的单次结果。

1.2 神经网络结构

1.2.1 结构参数化

神经网络在层数、宽度等需要人工设定的参数称为超参数。超参数可以用经验值设定初始值、取值范围，执行训练过程，并通过在测试集上的效果，选择最优的一组超参数。一般根据经验人工试验超参数，如果计算资源允许，可以用AutoML的方法自动搜索超参数。

1.2.2 残差网络

残差网络可以解决传统深度网络梯度弥散或梯度爆炸的问题^[12]。残差去掉相同的主体部分，突出微小的变化，减少了被跳过层的信息，使目标函数更容易拟合。我们使用神经网络去学习残差：

$$F(x) = H(x) - W_s x \quad (6)$$

其中： $H(x)$ 是残差结构与常规层整体的输出。如果 $F(x)$ 和 $H(x)$ 的维度相等， W_s 退化为常数1。残差网络跳过的层数 L_{res} 是需要调节的超参数。

1.2.3 损失函数

评估模型对应力预测的准确性，通常使用绝对误差率(MAPE)。但在神经网络的训练中，我们不能把这个函数当作目标。这个函数不可导，我们使用RMSE作为损失函数，从而可以使用梯度下降算法快速地训练神经网络。

1.2.4 神经元数量

神经网络的层数、宽度是重要的结构。层数过少，拟合能力差，甚至无法表达非线性的关系。另一方面，如果层数过多，网络学习难度大，拟合的效果也大大下降。在隐藏层中使用太少的神经元将导致欠拟合。相反，使用过多的神经元同样会导致一些问题，比如会导致过拟合^[13]。

当神经网络具有过多的节点时，训练集中包含的有限信息量不足以训练隐藏层中的所有神经元，因此就会导致过拟合。即使训练数据包含的信息量足够，隐藏层中过多的神经元会增加训练时间，从而难以达到预期的效果。显然，选择一个合适的隐藏层神经元数量是至关重要的。

1.2.5 早停

在小样本场景下，过拟合是很容易发生的。从训练最初开始，权重初始化为小值，这时，只会用到神经网络的线性部分，网络能力比较弱。随着训练时间的增长，会越来越多地用到网络的非线性部分，网络能力逐渐增强。这时观察验证集误差，如果持续增加的话，就可以提早停止训练。提前停止其实是另一种正则化方法，就是在训练集和验证集上，一次迭代之后计算各自的误差。绘制训练集误差、验证集误差与训练次数的曲线，选择验证集误差最低时的模型作为目标模型。

1.2.6 其他结构

隐藏层和输出层的神经元通过激活函数，将输入端的加权和转换成神经元输出值。在没有激活函数的作用下，无论怎么调整神经网络的参数，输出值仍为线性。加入更多的隐藏层本质上也是一样的。但是真实世界中的大多数系统是非线性的，若要模拟复杂系统，则必须借助非线性的激活函数。根据通用近似定理，神经网络至少需要一层隐藏层和足够的神经元，利用非线性的激活函数，便可以模拟任何复杂的连续函数。常用的Sigmoid、Tanh ReLU、Leaky-ReLU、ELU我们可以依次尝试。

遗忘单元指在深度学习网络的训练过程中，对于神经网络单元，按照一定的概率将其暂时从网络中丢弃。

正则化可以减少模型的熵，是减轻过拟合的有效方法。L1正则化将权重精确地推向零，从而更易成为

稀疏模型。L2正则化将惩罚权重参数而不使其稀疏，因为对于较小的权重，惩罚变为零。

1.3 迁移学习

数值模拟和机理模型，蕴含了大量的专家知识，这些知识如果能够迁移到神经网络中，有机会显著提升模型精度。迁移学习把一个场景的知识迁移到新的场景上。规则、公式、计算程序，如果用它们产生数据，这些数据就是Ds。

增量学习是指场景A训练好的神经网络模型，用场景B的数据继续训练，可以一定程度上实现小数据集模型训练。增量学习是自然语言、图像等领域内最常用的方法之一，其中场景A一般对应非常大的数据集，在建立应力-应变本构模型的过程中，我们也需要一个方式生成大量数据去训练场景A的模型。场景A的数据可以通过本实验中拟合的Arrhenius模型生成大量数据，或者建立一个可以适配多种合金的模型。然后用本场景的实验数据，继续训练该神经网络。这样，我们就完成了增量学习。此外，在数据充足的场景，也可以使用增量学习，继续提高小样本神经网络的模型精度。

不同合金、不同条件下的应力-应变本构模型，背后一定具有相同的规律，即使这些规律很复杂，人类不方便量化地理解，神经网络却可以最大程度地学习到这些量化的规律。通过对“学习过程”的学习，在新的数据集上，即使用小样本，也能迅速使新的模型达到很高的精度。元学习分类中的二阶梯度网络^[14]，其中，一阶网络执行每个小数据集的学习任务，二阶神经网络对一阶神经网络的参数求导，从而在新的极少样本上快速收敛。

迁移学习的方法还有很多，一般根据能够获取的数据情况、数据质量选择合适的方法。

2 试验与结论

2.1 材料

试验所用材料为定向凝固镍基高温合金DZ411，其标称成分如表1所示^[15]。铸件采用真空感应熔炼（VIM）和定向凝固（DS）技术生产，柱状晶晶粒生长方向与定向凝固方向基本一致，且与<001>晶体取向之间的角度差小于5°。在光学金相显微镜及场发射扫描电子显微镜下观察，合金微观组织主要由γ相基体

表1 DZ411合金化学成分表
Table 1 Chemical composition of DZ411

C	Cr	B	Co	W	Mo	Al	Ti	Ta	Ni	$w_B/\%$
0.1	14.0	0.01	10.0	3.50	1.30	3.40	5.20	3.10	余量	

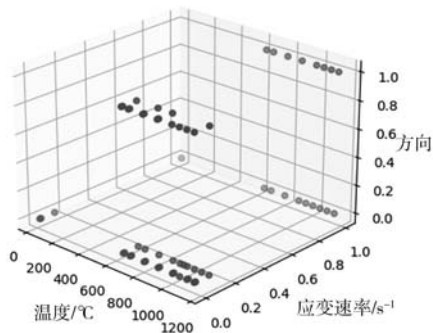
组织、立方状 γ' 强化相及少量MC及 $M_{23}C_6$ 碳化物构成。为保证合金成分均匀性，对铸件进行了均匀化热处理，热处理规程为 $1\ 300\ ^\circ\text{C} \times 3\ \text{h}/\text{AC}+1\ 080\ ^\circ\text{C} \times 6\ \text{h}/\text{AC}+870\ ^\circ\text{C} \times 20\ \text{h}/\text{AC}$ 。均匀化后的材料分别沿生长方向（ $\langle 001 \rangle$ 方向，轴向）及垂直生长方向（径向）取样，采用电火花线切割方式加工为 $\Phi 6\ \text{mm} \times 10\ \text{mm}$ 尺寸圆柱形试样，避免引入额外的塑性应变，并对两端进行简单打磨，降低粗糙度。

2.2 数据获取

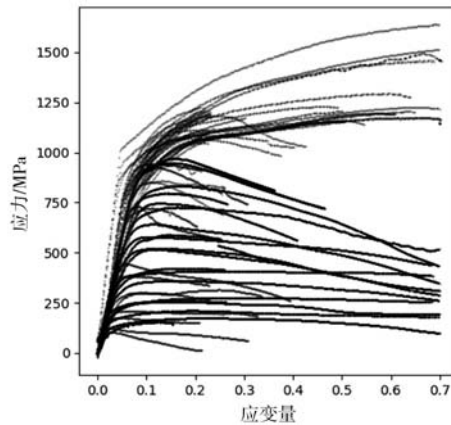
首先设计试验参数。选取试验参数范围：应变速率分别为0.001、0.01、0.1、 $1\ \text{s}^{-1}$ ，温度在室温（ $20\ ^\circ\text{C}$ ）、中低温（ $650\sim 900\ ^\circ\text{C}$ ）及高温（ $950\sim 1\ 170\ ^\circ\text{C}$ ）温度范

围内，方向为轴向及径向两个方向，应变 $0\sim 0.7$ ，选取测试集（温度，应变速率，方向）为（975，0.1，0）、（900，0.1，1）、（800，0.1，1），测试集不参与任何模型的训练、拟合。参考数据分布章节，选取试验参数。

将试验参数或者采集到的数据分别绘制为2维、3维的分布图表，对应图4a，发现分布不是正态分布，我们使用正态分布变换做数据映射。将不同条件下的应力-应变绘制为曲线，对应图4b，发现部分曲线呈现周期性波动，其中，应变是变形量与原始量的比值。为了训练Arrhenius模型我们将曲线平滑处理；训练神经网络的数据不需要平滑处理。特征列只有4个，不需要剔除特征。



(a) 分布表



(b) 应力-应变曲线

图4 试验参数及不同条件下的应力-应变曲线

Fig. 4 Experimental parameters and stress-strain curves under different conditions

根据参数表进行材料试验。材料试验为变形试验，在Gleeble-3500型热模拟试验机上开展，最终真实应变量为0.693。每个试样以 $10\ ^\circ\text{C}/\text{s}$ 速度加热至目标温度，并保温5 min。为尽量减少摩擦对应力状态造成的影响，在试样两端添加钽片进行润滑。通过计算机记录热压缩过程实时温度及应力、应变数据，并在试验结束后立即冷水淬火固定组织。

试验数据的记录格式为四元组（应变速率，温度，应变，方向，试验序号，试验内时间序号）。其中，试验序号、试验内时间序号，在特征预处理、特征增强中会使用到。将试验数据保存为CSV格式，共得到62个训练集、3个测试集。对于每一条应力-应变曲线，使用SMOTE数据增强的方法，合成训练数据，混入真实数据之中一同作为训练数据集。

2.3 模型训练

Arrhenius模型用标准的过程拟合，输入使用了共24个高温区、轴向数据。在Arrhenius模型拟合的过程

中遇到一些困难。第一，需要多次将数据点绘制为直线求斜率，而数据点经常只有2、3个，按照流程计算了斜率，但由于采样点过少，这个斜率一定包含不可控的误差，这在小样本情况下是无法解决的。第二，试验记录仪器无法精准定位过程中的应变，这导致需要保证应变相等的直线上，需要增加合成的数据点，或者丢弃这种数据，这种人为操作也会引入不可控的误差；反复试验直到采集到理想的数据，可以降低模型误差率。但是，增加试验会带来成本提高，也不是生产实践中合理的做法，所以不再追加试验的方法继续优化Arrhenius模型。

BP神经网络在不提供源代码、超参数、数据处理程序的情况下，一般很难精准复现其他论文中的效果。这是因为，数据处理方法灵活、超参数多、一些人工操作的步骤因人而异。BP神经网络是小样本神经网络的一部分，训练过程中我们也尽可能地优化。

小样本机器学习流程如图5所示，其中迁移模型和目标模型是结构相同的神经网络，该网络在BP神

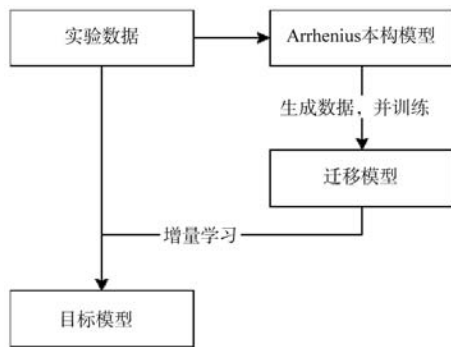


图5 小样本机器学习流程(增量学习)

Fig. 5 Few sample machine learning process (Incremental Learning)

神经网络的基础上, 增加残差单元。使用增量学习的方法, 将Arrhenius模型实现为Python程序, 在神经网络训练的过程中, 随机产生输入(应变速率, 温度, 应变变量), 令Arrhenius模型输出应力, 输出、输入整合为一行训练样本。通过反复的超参数调优, 生成与Arrhenius模型等价的迁移模型。由于Arrhenius可以无限地生成数据, 理论上, 迁移模型可以与Arrhenius达到一样的精度, 但是, 过分拟合Arrhenius并不是最终目标, 应当以增量学习后的效果作为目标, 多次试验迁移模型的最佳停止精度。我们通过试验选择在误差率略高于被迁移模型的时候停止迁移模型的训练。

损失函数使用RMSE, 训练过程中同时输出MAPE和RMSE, 试验中观察到两个指标的变化趋势是一致的, 说明损失函数的选择是有效的。

网络结构根据经验值设定初始值和搜索范围, 反复训练并观察模型性能, 在搜索范围内调节到最优超参数组合。其中, 输入层和输出层的节点数量很容易得到。输入层的神经元数量等于待处理数据中输入变量的数量, 输出层的神经元的数量等于与每个输入关联的输出的数量。我们通过试验选择最合适的层数, 但需要根据经验选取试验的范围。温度、应力、应变的关系不是线性关系, 也不是常见的多项式关系。对比 $900\text{ }^{\circ}\text{C}-0.001\text{ s}^{-1}$ 曲线和 $900\text{ }^{\circ}\text{C}-0.01\text{ s}^{-1}$ 可以发现, 在 $\dot{\epsilon}=0.1$ 的位置, 前者的斜率先升后降, 后者的斜率持续下降。这种非线性关系, 不仅需要多个隐含层, 而且需要激活函数, 增加网络的非线性拟合能力。

网络结构寻优过程中, 记录了一些主要超参数与模型性能的关系。表2中A代表激活层, 数字代表线性层的神经元数量, R代表残差单元的连接。观察训练过程中训练集损失和测试集损失的变化过程, 如果训练集损失不能下降, 可以依次调节更小的学习率、更简单的网络结构。但为了最终的误差率更低, 我们需要依次调节更复杂的网络结构。学习率、批大小等参数, 在其他参数固定的情况下, 需要逐个测试比较。表3证明训练集的数量一般是越多越好, 但增加数据的

边际收益降低。

如表2、3所示, 获得了单个影响因子对误差率的影响, 这是局部参数, 仍然需要找到全局的最优参数。采用循环优化的方法来调节最优参数。首先, 找到最优的网络结构, 固定网络结构, 继续调节学习率等参数。固定学习率等参数, 再次调节网络结构, 找到最优的结构, 然后再调节学习率等参数。如此反复, 直到找到这条线索上的最优参数组合。优化网络结构时, 发现两种结构的误差率都有优势, 沿着这两条线索分别去优化。这样, 就得到了全局最优参数。最终, 结合增量学习和超参数调整, 得到的小样本神经网络在测试集上的误差率为0.041。重要的超参数如表4所示, 网络结构如图6所示, 其中Linear是普通的神经网络层, 即线性层, in/out分别表示输入/输出的数量; ELU是指数线性激活函数^[6]; 从下面开始, 第一个线性层的输出, 叠加到最后一个线性层之前, 这一部分是残差网络结构。随着原始数据、数据增强等环节的变化, 网络结构超参数也需要重新搜索, 由于单次训练寻优比较耗时, 并未使用自动超参数搜索, 而是

表2 网络结构对误差率的影响
Table 2 Influence of network structure on MAPE

网络结构	MAPE
4-4-A-4-A-1	0.582
4-4-A-1	0.313
4-12-A-4-A-1	0.227
4-4-A-2-A-1	0.213
4-8-A-4-A-1	0.138
4-R-8-A-4-A-R-1	0.126

表3 训练集数量对误差的影响
Table 3 Influence of number of training sets on MAPE

训练集数量	MAPE
12	0.112
16	0.077
62	0.053

表4 主要参数的全局最优组合
Table 4 Global optimal combination of the main parameters

参数名称	取值
网络结构	4-R-8-A-4-A-R-1
权重衰减	1.00E-06
训练轮次	5 849
样本采样率	1
早停	1 600
学习率	0.000 4
批大小	128

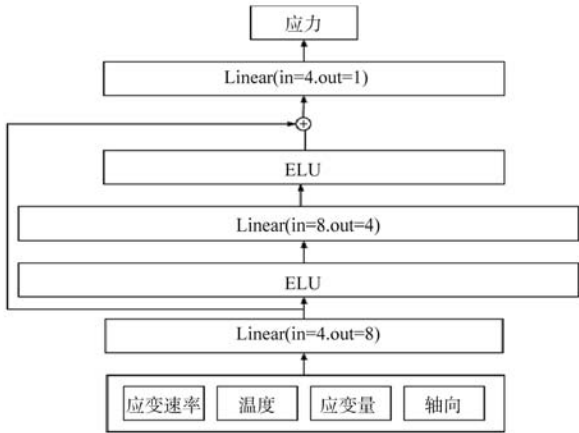


图6 优化后的小样本神经网络结构
Fig. 6 Optimized few sample neural network structure

依靠经验调整超参数的范围，最终也能得到精度较高的模型。

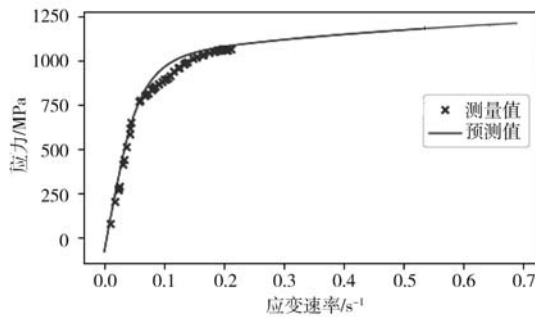
2.4 试验结果和分析

模型的比较如表5所示。Arrhenius模型在高温、轴

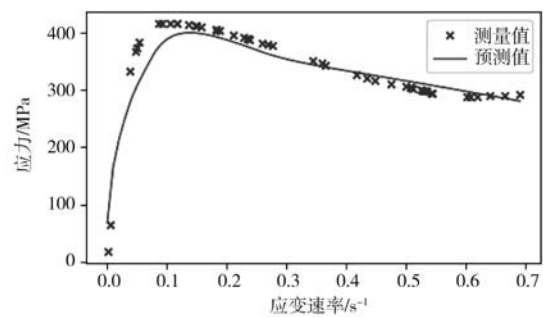
向、应变率0.21以内测试集上的误差率是0.375，在其他温度区间、径向方向上，误差率增加到0.607。小样本机器学习的误差率是0.041，测试范围是全部试验区间，即高温、低温、径向、轴向的情况都做了测试，综合性能优势明显。小样本神经网络允许方向作为输入，并且能够自动适配低温、高温区的特性，这是Arrhenius模型无法具备的能力。图7a-c绘制了目标模型在未知数据上的拟合结果，从曲线形态上看，试验曲线与拟合曲线有交叉，说明模型回归充分；拟合曲线比试验曲线曲率变化更简单，说明模型为了提升泛化能力，简化了模型参数，牺牲了一定的准确性。由于试验数据采集是有误差的，过度拟合并非最优解，这种模型参数简化是合理的。

表5 模型性能比较
Table 5 Performance comparison of different models

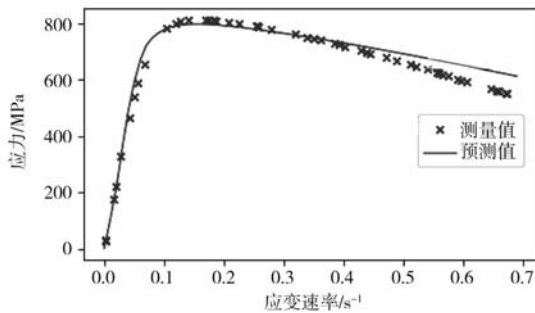
方法	区间误差率 (高温区, 变形量<0.21)	整体误差率
Arrhenius	0.375	0.607
小样本机器学习	0.041	0.041



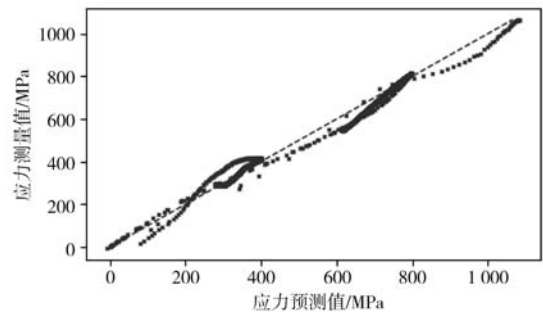
(a) 温度=700 °C, 应变速率=0.1 s⁻¹, 方向=径向



(b) 温度=1050 °C, 应变速率=0.01 s⁻¹, 方向=轴向



(c) 温度=975 °C, 应变速率=1.0 s⁻¹, 方向=轴向



(d) R²=0.97

图7 小样本机器学习应力-应变本构模型对未知数据的预测结果

Fig. 7 Prediction of unknown data by proposed model

3 结束语

本研究在DZ411高温合金变形试验中，采集了共65个试验数据，其中3组为测试集。分别构建了Arrhenius模型和小样本机器学习模型。

Arrhenius模型的拟合对样本采集点的要求很高、对样本采集的误差很敏感，需要反复实验、依靠人工经验去除有误差的数据，才能达到理想的性能。

小样本机器学习，在既有的数据下，无需重复及

追加试验。循环执行训练流程,结合数据增强、网络结构、迁移学习的全面超参数优化,最终得到了最优的超参数组合,构建的小样本神经网络模型应力-应变本构模型,在未知数据上的预测误差率为0.041,精度显著高于同等试验条件下的Arrhenius模型。

在更广泛的铸造生产实践中,经常遇到生产条件不容易复现、试验成本高的情况。在这种只有小样本

数据的情况下,构建高精度的模型是一个挑战。公式类模型对数据质量依赖高,可能需要依赖较多试验,输入、输出参数固定,往往不是最佳选择。本研究提出的小样本机器学习,可以充分利用现有的公式,不仅对小样本拟合精度更高,并且不受公式限制允许输入更多参数,在未来的铸造数值模拟中,一定能发挥重要的作用。

参考文献:

- [1] 佚名. 高温合金为航空、航天、能源事业保驾护航 [J]. 材料保护, 2022, 55 (3): 32.
- [2] 杨屹, 蒋玉明, 刘力菱, 等. 铸件凝固过程中热应力场及热裂的数值模拟研究分析 [J]. 铸造技术, 2000 (2): 36-39.
- [3] 钟万龠, 姚伟岸, 郑长良. Reissner板弯曲与平面偶应力模拟 [J]. 大连理工大学学报, 2002, 42 (5): 519-521.
- [4] 陶攀, 李怀学, 黄柏颖, 等. 激光选区熔化AlSi10Mg合金悬臂梁残余变形的数值模拟 [J]. 特种铸造及有色合金, 2018, 38 (5): 493-497.
- [5] 朱浩, 朱亮, 陈剑虹. 铝合金在三种应力状态下的力学性能研究及断口分析 [J]. 兰州理工大学学报, 2006, 32 (6): 28-31.
- [6] ZHONG Jiangwei, SUN Changbo, WU Jiantao, et al. Study on high temperature mechanical behavior and microstructure evolution of Ni₃Al-based superalloy JG4246A [J]. Journal of Materials Research and Technology, 2020, 9: 6745-6758.
- [7] ZHANG B. Sample-centric feature generation for semi-supervised few-shot learning [J]. IEEE Transactions on Image Processing. 2022 (31): 2309-2320.
- [8] HU Y, GRIPON V, PATEUX S. Leveraging the feature distribution in transfer-based few-shot learning [C]// Artificial Neural Networks and Machine Learning. In International Conference on Artificial Neural Networks, 2021: 487-499.
- [9] FRAZIER, PETER I. A tutorial on Bayesian optimization [J]. Ar Xiv: 2018, 1807: 02811.
- [10] CAO Y, KO S W. Fusing separated representation into an auto-encoder for magnetic materials outlier detection [J]. Systems Science & Control Engineering, 2022 (10): 181-191.
- [11] CHENG Z Y, ZOU C M, DONG J W. Outlier detection using isolation forest and local outlier factor [C]// In Proceedings of the conference on research in adaptive and convergent systems, 2019: 161-168.
- [12] JHA D, WARD L, YANG Z, et al. IRNet: a general purpose deep residual regression framework for materials discovery [C]// In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, 6: 2385-2393.
- [13] YANN L C, BENGIO Y, HINTON G, et al. Deep learning [J]. Nature, 2015, 521: 436-444.
- [14] FINN, CHELSEA, PIETER A, et al. Model-agnostic meta-learning for fast adaptation of deep networks [C]// International conference on machine learning, 2017: 1126-1135.
- [15] 《高温合金手册》编写组. 高温合金手册 [M]. 北京: 冶金工业出版社, 1972.
- [16] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs) [C]// ICLR, 2016, Ar Xiv: 1511: 07289.

Construction of Stress Strain Constitutive Model of Ni-Based Superalloy Based on Few Sample Machine Learning

YIN Rong¹, SUN Jia-yan², XU Qing-yan²

(1. Xiaomi AI Lab, Beijing 100102, China; 2. Key laboratory for Advanced Materials Processing Technology (MOE), School of Materials Science and Engineering, Tsinghua University, Beijing 100083, China)

Abstract:

In the process of constructing the mechanical constitutive model of a Ni-based superalloy, this paper uses few sample machine learning which includes data enhancement, network structure optimization, transfer learning, and other methods to construct a few sample neural network to reduce the count of experiments. The tests show that the proposed method needs fewer data and has higher accuracy than standard BP neural network and phenomenological constitutive model.

Key words:

superalloy; stress-strain constitutive model; few sample; machine learning