

引文格式:袁睿豪,廖玮杰,唐斌,等.数据驱动的航空发动机材料设计研究进展[J].航空制造技术,2021,64(18):22-30.

YUAN Ruihao, LIAO Weijie, TANG Bin, et al. Progress of data-driven aero-engine materials design[J]. Aeronautical Manufacturing Technology, 2021, 64(18): 22-30.

# 数据驱动的航空发动机材料 设计研究进展\*

袁睿豪,廖玮杰,唐斌,樊江昆,王军,寇宏超,李金山

(西北工业大学凝固技术国家重点实验室,西安710072)

[摘要] 数据驱动的材料设计模式已经广泛应用于材料科学研究中,以加速新材料从研发到应用的进程。机器学习技术能够挖掘数据中存在的潜在模式或规律,是数据驱动材料设计模式的研究关键和热点。其中,基于主动学习的策略已经成功指导了多种新材料的开发。首先,介绍了主动学习的研究思路、构成要素及每个环节常用的方法;其次,以航空发动机材料如高温合金、钛基合金、复合材料、热障涂层等为例,介绍了机器学习在其中的具体应用和取得的效果;最后,针对航空发动机材料的恶劣服役环境,展望了机器学习应用于航空发动机材料面临的挑战,并提出了可能的解决方案。

关键词:材料基因工程;机器学习;主动学习;航空发动机材料;高温合金

DOI: 10.16080/j.issn1671-833x.2021.18.022



袁睿豪

副教授,博士,主要研究方向为材料信息学。

航空发动机是现代工业技术的明珠,航空发动机材料是高性能航空发动机的基础。伴随着航空发动机的发展,其主要部件所采用的材料也在不断地更新换代。例如,从早期一、二代发动机主要使用金属结构材料如高温合金,到第四代发动机中复合材料的引入<sup>[1]</sup>。近年来,新一代航空发动机对推重比、压气机增压比、涡轮前入口温度等性能指标提出了更高的要求。因此,设计和开发新型的航空发动机材料,满足在更高温度、更高压力、更高速度等极端环境下的长期服役要求,具有重要的理论意义和工程应用价值<sup>[2]</sup>。但是,复杂的成分构成和加工工艺使得航空发动机材料的设计和开发耗时较长,成本较高。以传统的高温合金为例,多种合

金元素可用于合金化,导致了巨大的未知成分空间;另一方面,为了优化合金的微观组织结构,需要同时调控热处理温度、时间、变形量等多种加工工艺参数,产生了巨大的未知工艺参数空间<sup>[3]</sup>。针对这些问题,传统的材料研发方法如试错法和经验法难以实现材料的快速筛选、设计和开发,亟须发展新的研究手段,提升航空发动机材料的研发效率。

材料基因工程是近年来新兴的材料研发新范式,通过融合高通量试验、计算和大数据技术,实现材料研发时间和成本的同时减半<sup>[4]</sup>。其中,基于高通量试验和计算产生的数据,利用大数据技术,挖掘材料数据中潜在的模式或者物理规律,加速新材料的设计与开发,是近年来的研究热

\* 基金项目:国家重大科技专项(MJ-2018-G-48)。

点。与传统的计算和理论模拟不同,数据技术如机器学习等仅需要从数据出发,构建成分、工艺、组织等和目标性能之间的映射关系,实现对所需性能的正向预测,或者基于构建的预测模型,逆向指导最佳成分或工艺参数的选择<sup>[5]</sup>。这一研究方法在一定程度上避免了理论模型难以处理复杂材料体系的难题,例如,相图计算方法需要大量的热力学数据,对于多元体系的计算精度难以保证;第一性原理计算能够保证计算精度,但是计算规模通常是几百个原子,限制了在复杂多元体系中的应用<sup>[6]</sup>。值得注意的是,基于机器学习的材料开发是从数据出发的,意味着这一研究方法不受限于特定尺度,可以实现微观、介观、宏观的多尺度建模。

最近,数据驱动的研究思路也逐渐被用于航空发动机材料的研究中,以提升材料的研发效率,预测极端条件下的服役性能等。本论文主要从以下3个角度进行讨论:首先,以最近广泛应用的主动学习为例,详细讨论如何进行数据库构建、机器学习建模和模型评估、优化算法指导材料选择以及试验或计算的验证;其次,以航空发动机材料,即高温合金、钛合金、复合材料、热障涂层为例,介绍了机器学习在其中的具体应用;最后,对机器学习在材料中的应用进行总结,针对航空发动机服役环境的复杂性,讨论了机器学习在航空发动机材料的研究中所面临的难题,并提出了可能的解决思路。

## 数据驱动的材料研究进展及基本思路

目前,基于机器学习的数据驱动材料研究主要集中于以下3个方面:(1)通过构建成分、组织、结构与性质、性能之间的定量预测模型,加速新材料的开发;(2)基于大量的数据,提取潜在的规律,获得新的物理知识;(3)通过融合已有的成熟理论

模型,进一步加速数据模型的性能优化效率。例如,通过直接构建高温合金成分与 $\gamma'$ 相溶解温度的数据模型,实现对具有高 $\gamma'$ 相溶解温度合金成分的快速筛选<sup>[7]</sup>;构建材料的描述符,通过特征工程,确定影响材料性能的关键基因,指导新材料的设计<sup>[8]</sup>;通过将已有的理论与数据模型结合,对未知材料空间进行预先筛选,实现高性能铁电材料的快速开发等<sup>[9]</sup>。其中,基于主动学习的材料开发策略在不同的材料体系中均获得了成功<sup>[10]</sup>。因此,本文将主动学习为例,简要介绍数据驱动材料设计的思路和关键步骤,主要包括以下5个部分:材料数据库的特点及构建;材料描述符的构建及筛选;机器学习模型的构建和评估;优化试验设计;试验/计算验证及反馈。

### 1 主动学习简介

与传统的基于机器学习的材料设计思路相比,主动学习包括优化试验设计和数据反馈环节,这一环节能够在迭代的过程中快速提升机器学习模型的性能,进而加快材料研发效率。如图1所示<sup>[10]</sup>,首先,基于文献或实验室累计数据,构建所需的数据库,并利用数据算法,对数据进行清洗、降维等处理,以满足后续机器学习建模的需求;其次,根据对所研究材料的理解,构建材料描述符,增强后续机器学习模型的预测能力;再次,不同的机器学习算法均能够用于数据模型的构建,因此,需要利用标

准误差等判据对数据模型的性能进行评估;相对于未知数据,已有数据体量通常较小,导致预测过程存在不确定性,基于预测值和伴随的不确定性,结合优化算法推荐最具期望的新材料进行验证;最后,基于算法的推荐,进行样品的制备、测试与表征(或者第一性原理计算),并将数据反馈到初始数据集,进行下一轮迭代,直至找到满足要求的目标材料<sup>[10]</sup>。在主动学习中,迭代停止准则可以基于以下3个方面考虑:材料已经满足目标性能;优化算法的提升指标已经收敛;预算难以满足更多次迭代。

### 2 数据库的构建

高质量的数据库是保证数据模型的预测精度和外推能力的基础。对于主动学习,数据库需要包括输入变量和输出变量,输入数据通常包括材料的元素含量、种类、晶体结构、加工工艺、微观结构等,输出数据包括材料的目标属性如力学性能、物理性能、化学性能等。在实际应用中,难以快速获得高质量的试验数据和理论模拟数据,需要借助文献中已经发表的数据或者已开放的数据库。对于文献数据,可以采用文本挖掘结合自然语言处理实现对数据的快速提取<sup>[11-12]</sup>。另外,目前存在多个开源数据库,例如Materials Project<sup>[13]</sup>、AFLOW<sup>[14]</sup>以及无机晶体结构数据库(ICSD)<sup>[15]</sup>等,包含数十万条数据,能够为模型提供大量的数据进行训练。值得注意的是,由于试验条件

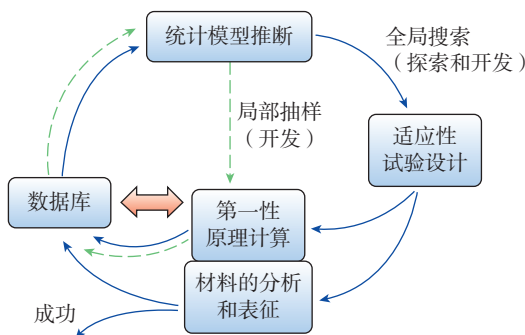


图1 基于主动学习的材料设计

Fig.1 Materials design based on active learning

或者操作方式不一致,文献中的数据可能存在较大的不确定性,即对同种材料同种工艺,可能会产生偏差很大的结果,这需要在在使用过程中仔细甄别。同时,已有的开源数据库通常基于理论模拟计算,和试验测试数据之间不可避免地存在偏差,如何将理论模拟数据用于指导试验设计,需要研究人员进行合理的考量。基于文献或者开源数据库建立的数据集,可能存在数据缺失或者重复的问题,可以利用机器学习算法对数据进行清洗、补全等处理,保证后续高性能机器学习模型的建立。

### 3 材料描述符的构建与筛选

材料描述符需要具有以下性质:相对于目标性能更容易获得、维度适中、具有可解释性。优异的描述符在提升数据模型性能的同时,能够增强模型的可解释性。材料描述符的构建可以采用手动和自动两种方式。前者通常包括以下3类:(1)材料的物理特性如原子半径、电负性等;(2)材料晶体结构的坐标等;(3)对结构或者成分的二进制编码表示(One-hot)<sup>[7,16-17]</sup>。后者可以采用神经网络模型等对已有数据如组织结构图片进行建模,提取能够代表图片信息的关键潜在变量,即描述符,进行图片的重构和后续优化<sup>[18-19]</sup>。由于描述符的数值在数量级上会存在较大差别,可能导致模型偏向于某一特定描述符。因此,在建立数据模型之前,通常需要对描述符进行归一化处理,即将描述符映射到某一特定数值区间内,常用的归一化公式有:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \text{ 或 } x^* = \frac{x - \bar{x}}{x_{\max} - x_{\min}} \quad (1)$$

分别可以将数据映射到[0,1]和[-1,1]区间,其中 $x_{\max}$ 和 $x_{\min}$ 分别表示某一描述符数据的最大值和最小值; $\bar{x}$ 为该描述符数据的均值<sup>[20]</sup>。

描述符维度过高一方面会降低

模型的训练效率,另一方面,不同描述符可能代表着类似的信息。为此,需要对描述符进行预先筛选,提取与目标性能最相关的描述符。评估描述符本身与目标性能之间的相关性的算法有多种,如皮尔逊相关系数(Pearson correlation coefficient, PCC),这一算法可以表征两个变量之间的线性相关性,但对于一些非线性的关系则难以捕捉<sup>[21]</sup>;最大互信息系数(Maximal information coefficient, MIC)能够评估描述符间的复杂函数关系,如抛物线、三角函数等非线性关系。常用的描述符相关性表征算法和特点如表1所示<sup>[22]</sup>。除了相关性分析,还有重要性排序算法如Lasso、随机森林、决策树等,能够根据描述符对目标性能的影响程度进行重要性排序<sup>[27-28]</sup>。经过筛选的描述符,可以进一步通过最佳子集选择、遗传算法等,确定最佳的描述符组合。

### 4 机器学习模型的构建与评估

基于构建的数据库,可以用多种机器学习算法训练代理模型,例如,线性模型、多项式模型、支持向量机、神经网络等。线性模型和多项式模型是基于最小二乘法,通过将均方误差最小化确定模型的参数。这一算法形式简单、容易解释,但效果通常难以满足要求。对于具有复杂模式的数据,需要更加复杂的算法如支持向量机和神经网络等。支持向量机在处理小数据量问题时展现了优异的性能,基本思想是利用核函数将数据映射到更高维度的空间,通过最大化不同类之间的超平面,实现分类效果,如图2所示,其中少数支持向量决定了最终结果,这使得该模型不仅算法简单,而且可以抓住关键样本,去除冗余样本对分类的影响<sup>[29]</sup>。但是,支持向量机算法在应对大规模样本时会耗费大量的机器内存和运算

表1 相关性算法与特性

Table 1 Correlation algorithms and characteristics

相关性算法	适用范围	是否标准化
MIC	线性、非线性数据	是
Pearson	线性数据	是
Spearman <sup>[23]</sup>	线性、简单单调非线性数据	是
Kendall <sup>[24]</sup>	线性、简单单调非线性数据	是
阈值相关 <sup>[25]</sup>	线性、非线性数据	是
距离相关 <sup>[26]</sup>	线性、非线性数据	是

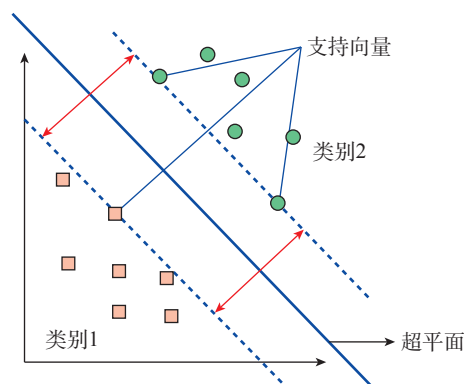


图2 支持向量机算法的二维平面表示

Fig.2 Two-dimensional plane representation of support vector machine algorithm

时间从而难以实施,对于多分类问题也需要结合其他算法或组合多个二分类支持向量机解决。神经网络模型理论上可以映射任何复杂的非线性关系,但一般需要很多数据训练模型,同时,过多的超参数使得模型训练耗时长、成本高,并且模型通常难以解释。

在数据模型构建之后,为了避免过拟合和欠拟合,需要对模型性能进行评估。通常采用交叉验证的方法,将初始数据集划分为训练集和测试集,计算模型的误差或者精度。例如十折交叉验证是将训练数据集等分为10个子集,每次选用其中9个子集作为训练集训练模型,剩余的一个子集用作测试集测试模型性能,依次迭代10次,直至每个子集都曾被用作测试集。由于十折交叉验证利用训练集的子集进行了测试验证,因此能够避免预测模型的过拟合,更可靠地评估模型的泛化能力。对于分类模型,常用的性能评估指标有准确率、错误率、ROC曲线等。对于回归模型,常用的指标有平均绝对误差(Mean absolute error, MAE)、均方误差(Mean squared error, MSE)、决定系数( $R^2$ )等,表达式分别如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (4)$$

式中,  $y_i$  表示目标性能的实际值;  $\bar{y}_i$  表示实际值的均值;  $\hat{y}_i$  表示模型的预测值;  $N$  为数据集材料的个数。

## 5 优化试验设计

未知材料空间巨大,无法对所

有候选材料组分或者工艺参数进行验证。因此,需要借助优化算法,推荐最有可能提升材料性能的成分或者工艺参数组合,最大程度减少试验验证的次数<sup>[10]</sup>。常用的优化算法基于以下两种策略的平衡,即探索(Exploration)和开发(Exploitation)。探索的策略主要考虑选取未知材料分布中预测值不确定性最大的成分点,反馈到初始数据集,最大程度矫正数据模型对预测不准确的数据点的评估,但是,所推荐成分的性能通常难以满足要求。开发的策略基于对未知材料的预测,直接选择具有最大预测值的材料进行试验验证,虽然能够发现高性能材料,但可能会限于局部极值。因此,需要能够同时平衡开发和探索的策略,即全局优化算法(Efficient global optimization),可以同时考虑预测值和伴随的不确定性,在保证推荐高性能材料的同时,能够跳出局部极值,最终提升新材料的开发效率。

## 6 验证和数据反馈

主动学习中的关键一步便是数据的反馈,因此,基于优化算法推荐的成分或者工艺,需要进行试验或者模拟验证,将数据反馈到初始数据集,快速提升数据模型的性能,提升迭代的效率。目前,所用的验证方法主要为试验。但是,对于某些极端条件或者电子尺度的性能,试验表征难以实现,可以选择理论模拟的手段进行验证。例如基于密度泛函理论(Density functional theory, DFT)可以计算晶体在OK下的各种性质如弹性模量、带隙等<sup>[30]</sup>;基于热力学理论的CALPHAD( Calculation of phase diagram)计算可以模拟和预测材料的相图<sup>[31]</sup>;分子动力学(MD)和有限元模拟(FEM)可以分别在微观尺度和介观尺度上预测材料的相变和力学行为<sup>[32-33]</sup>。

## 数据驱动方法在航空发动机材料中的应用

航空发动机材料的恶劣服役环境给材料的研发带来了巨大的挑战,新材料从开发到应用可能要经过数十年的时间。本部分将以典型的航空发动机材料为例,如高温合金、钛基合金、陶瓷基复合材料、热障涂层等,介绍基于机器学习的数据驱动方法在其中的应用及最新进展。

### 1 高温合金

针对高温铁基奥氏体不锈钢的抗蠕变性能, Shin等<sup>[34]</sup>将热力学计算的高温下相体积分、相元素含量等作为输入变量建立模型,发现这些基于热力学计算的参量对蠕变性能的数据模型有较大影响。之后,该团队结合热力学参数和机器学习预测了9%~12% Cr铁素体/马氏体钢在不同温度下的屈服强度,发现在较低温度下模型的 $R^2$ 可以达到0.9以上,但是,随着温度升高,模型的准确率迅速下降,这可能是高温下数据缺乏和变形机制的改变所导致的<sup>[16]</sup>。在机器学习预测钴基或镍基高温合金性能的研究中, Yu等<sup>[7,35]</sup>采用合金成分,元素性质和时效温度、时间等热处理工艺参数作为描述符,成功预测了高温合金中 $\gamma'$ 相的溶解温度和体积分等; Zou等<sup>[36]</sup>采用多重扩散法(Diffusion-multiple approach)获得了1375个具有 $\gamma/\gamma'$ 两相区的材料成分和 $\gamma'$ 相的体积分,利用建立的数据库,通过随机森林、深度神经网络等算法构建了能够准确预测高温合金 $\gamma'$ 相体积分的机器学习模型。此外, Tamura等<sup>[37]</sup>利用机器学习建立了气体雾化过程参数与晶粒尺寸的关系,所得结果与Lubanska提出的公式中的趋势一致,并以此优化了Ni-Co基高温合金粉末冶金过程的工艺参数。镍基高温合金中拓扑密堆相(TCP)的存在会消耗强化元素并促进裂纹的产生,针对这一问题,

Qin 等<sup>[38]</sup>基于高通量试验获得的数据库,以合金成分和元素的混合焓、价电子浓度等物理参数为描述符,利用机器学习预测了合金中拓扑密堆相的存在,可以指导合金成分的快速筛选。上述利用材料微观结构的研究都采用了人工识别的方法,而在另外一项研究中, Khatavkar 等<sup>[18]</sup>利用图像识别技术提取材料微观结构信息作为描述符,如图 3 所示<sup>[18]</sup>。首先将 CMSX-6 和 DZ125 两种镍基高温合金的原始扫描电镜图像(图 3(a)和(d))进行阈值处理,产生只有黑白两种颜色的图像,如图 3(b)和(e)所示;之后通过图像识别技术提取图片信息如图 3(c)和(f)所示,作者将这种描述符用于机器学习并实现了对镍基高温合金硬度的精确预测。

## 2 钛基合金

人工神经网络等机器学习技术已广泛应用于钛基材料的研究中,如对 Ti-6Al-4V 合金的流动应力<sup>[39]</sup>、微观组织演化<sup>[40]</sup>、机械性能<sup>[41]</sup>与工艺参数<sup>[42]</sup>的预测等。Arisoy 等<sup>[43]</sup>利用随机森林模型研究了切削速度、刀具几何形状等工艺参数对合金显微组织和性能的影响,并通过遗传算法,根据所需的组织和硬度选择合适的加工工艺。Harsha 等<sup>[44]</sup>利用神经网络算法研究了切削速度、给进量、切削时间等输入参数与刀具磨损、钛合金表面粗糙度等输出响应之间的关系。另外,有研究通过机器学习预测了 Ti-6Al-4V 合金切削过程中的无颤振加工条件,并提高了工件表面的质量和刀具寿命<sup>[45]</sup>。对于其他的钛基合金,有学者利用神经网络模型,预测了纳米 B<sub>4</sub>C 粒子增强型 Ti-6Al-4V 合金磨损性能<sup>[46]</sup>;根据钛合金成分和热处理温度预测合金中的  $\alpha$  相和  $\beta$  相的体积分数<sup>[47]</sup>;建立了 Ti-2Al-9.2Mo-2Fe 合金应变变量、应变率、温度和流变应力之间的预测模型,优化了热加工条件<sup>[48]</sup>;以钛合金相组成、氧化温度、氧化时

间、氧气和水蒸气含量作为描述符,通过梯度提升树、随机森林和 k-近邻 3 种机器学习模型预测合金高温氧化的抛物线速率常数,其中梯度提升树模型的  $R^2$  可以达到 0.92,这是一种基于梯度下降法和决策树的集成算法,泛化能力较强,可以灵活处理各种类型的数据,包括连续值和离散值<sup>[49]</sup>;基于扩散复合技术产生的多组分变化数据库,通过 BP 神经网络建立了新型钛合金(Ti-3Al-2Nb-1.2V-1Zr-1Sn-xCr-yMo)的成分-显微组织-性能之间的模型,并成功设计出性能优异的钛合金<sup>[50]</sup>。

## 3 复合材料

复合材料因其优异的耐高温、抗

氧化、高比强度、抗蠕变等性能,是新一代航空发动机备受青睐的新型材料,包括陶瓷基复合材料、碳/碳复合材料、金属基复合材料等<sup>[51]</sup>。神经网络模型常应用于复合材料性能和行为的预测,如图 4 所示<sup>[52]</sup>, Pramod 等<sup>[52]</sup>建立了预测 Al<sub>2</sub>O<sub>3</sub> 颗粒增强 Al2075 复合材料磨损性能的人工神经网络模型,以密度、施加载荷、滑动距离和颗粒的质量分数作为描述符预测表面磨损高度,其中隐藏层的作用是把输入数据的特征抽象到另一个维度空间,使得更抽象化的特征能够更好地映射模型输出。SiC<sub>f</sub>-BN/SiC 复合材料中的纤维分布和基体形成层状结构,这种结构使得合金的

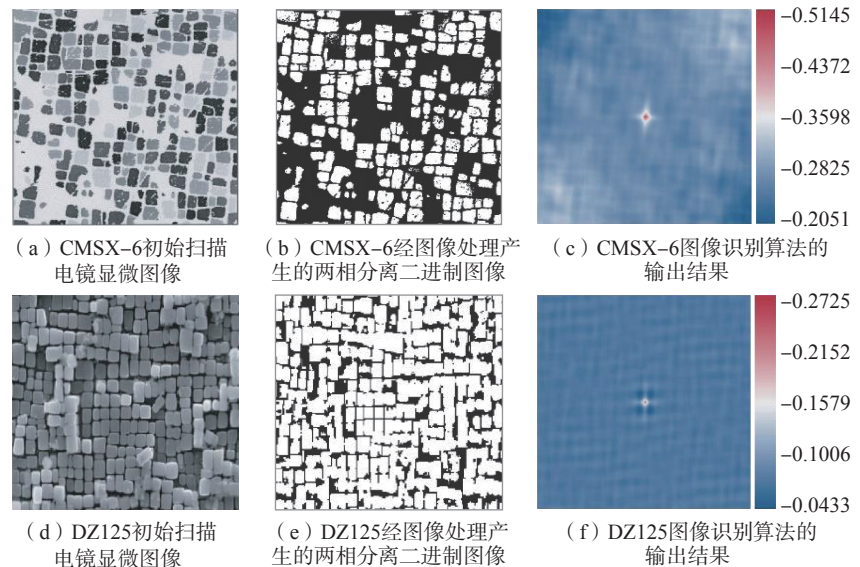


图 3 通过图像识别技术得到镍基高温合金微观结构信息

Fig.3 Microstructure information of Ni-based superalloys obtained through image recognition technology

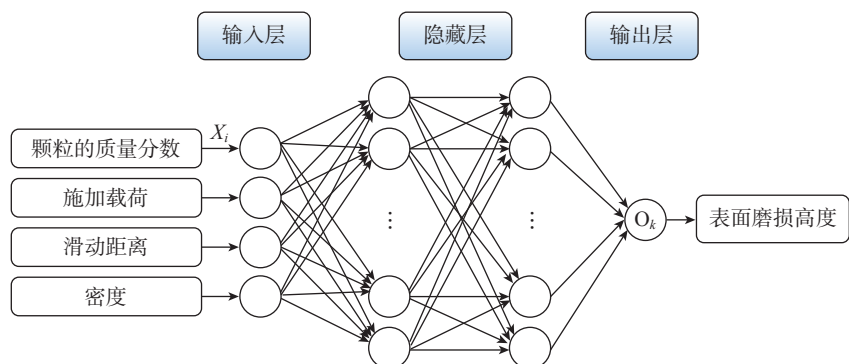


图 4 神经网络模型预测复合材料的磨损性能

Fig.4 Neural network model used to predict the wear performance of composite materials

陶瓷基体在横向载荷下产生的连续性受损成为引发失效事件的关键因素。针对这一现象, Patel 等<sup>[53]</sup>通过神经网络模型量化连续纤维增强陶瓷基复合材料在相关尺度下的损伤, 预测出最能抵抗基体损伤的纤维微观结构。Canakci 等<sup>[54]</sup>以时效温度、时间等热处理工艺参数、 $B_4C$  增强颗粒的尺寸和体积分数作为输入, 硬度、抗拉强度、屈服强度和弹性模量作为输出, 建立了预测 Al-Cu-Mg 基复合材料性能的神经网络模型(错误率仅有 2%), 发现热处理工艺很大程度上决定了材料性能。随着复合材料制造技术的提升, 分级、多孔、多维等复杂结构大大扩展了复合材料的设计空间, 也使得结构优化变得至关重要, 机器学习正是一种选择最佳结构的有效手段, 从而发挥复合材料的最大潜力。

#### 4 热障涂层

晶格热导率决定了非金属材料的导热能力, 是设计热障涂层的重要参数。Chen 等<sup>[55]</sup>通过基于随机森林的特征选择建立了高斯回归模型预测晶格热导率, 在 19 个样本的测试集里达到 0.93 ( $R^2$ ) 的准确性, 但由于数据集较小, 存在过拟合的风险。为了提高泛化能力, Juneja 等<sup>[56]</sup>结合通过高通量计算得到的描述符, 利用包含 2162 种材料的数据集进行机器学习, 得到了高鲁棒性的预测模型, 实现了对晶格热导率的准确预测。Loftis 等<sup>[57]</sup>关注了符号回归、多层感知器的深度神经网络、随机森林回归 3 种模型在 347 个样本的数据集上对晶格热导率的预测能力和不同特点, 随机森林通过集成学习降低了过拟合的风险; 多层感知器的深度神经网络能够从数据集中准确发现非线性关系; 符号回归虽然计算量大、耗时长, 但会产生具有物理意义的公式, 图 5 所示为符号回归得到的公式和预测结果<sup>[57]</sup>, 公式中的符号都为数据集中包含的描述符, 这对

理解描述符的作用和设计具有高晶格热导率的涂层材料有很大的价值。描述符的数据可以来源于各种测量仪器, 有学者利用神经网络模型学习红外成像数据, 模拟热响应与热障涂层厚度的关系<sup>[58]</sup>; 通过 BP 神经网络优化超声反射系数谱特征参数, 并利用高斯回归算法预测热障涂层的孔隙率, 其误差小于 5.3%<sup>[59]</sup>; 通过多元线性回归、BP 神经网络、支持向量机 3 种模型, 结合涂层的太赫兹光谱特征预测材料的微观结构(孔隙率、孔隙裂纹比、孔径等)<sup>[60]</sup>。数值型、图案型、图表型等不同类型的都可以通过特征工程转化为可用于机器学习的数据, 从而帮助理解机器学习模型和分析原始数据与目标性能的关系。

#### 5 分析与讨论

通过机器学习模型预测材料力学性能或断裂、氧化等失效行为的关键在于寻找合适的描述符, 从而提高预测模型的准确性和可解释性, 因此, 如何收集和选择描述符数据是材料研究与机器学习结合的重点之一。对于高温合金, 材料的微观组织结构是决定其性能的重要因素, 但这种

信息往往需要研究人员或专家从材料的显微图像中提取, 过于依赖科研人员的个人知识与经验。研发适用于材料科学的计算机图像识别技术, 可以从文献资料里大量的显微图像中快速有效地提取关键信息, 甚至得到一些研究人员难以直接观察到的结构信息。高温钛合金常用于航空航天发动机复杂结构件, 对性能要求较高的同时也提高了加工工艺的难度, 如何采集各种工艺参数及其对工件性能或结构变化的响应信息, 并将其转化为可用于机器学习的数据是应用这一技术设计工艺参数的关键。另一方面, 钛合金的组织与性能对热处理温度、时间等工艺参数敏感, 例如合金在  $\beta$  转变温度以上退火时, 晶粒会迅速长大, 从而降低材料的力学性能, 因此, 这要求机器学习模型能够准确预测材料性能并通过逆向设计推荐合适的参数。在复合材料领域, 机器学习对界面的研究还比较少见, 而这一因素对复合材料的物理、化学等性能有重要的影响, 界面的结合状态和强度, 以及界面的传递效应、阻断效应、不连续效应等都可以应用机器学习进行研究。降低热导

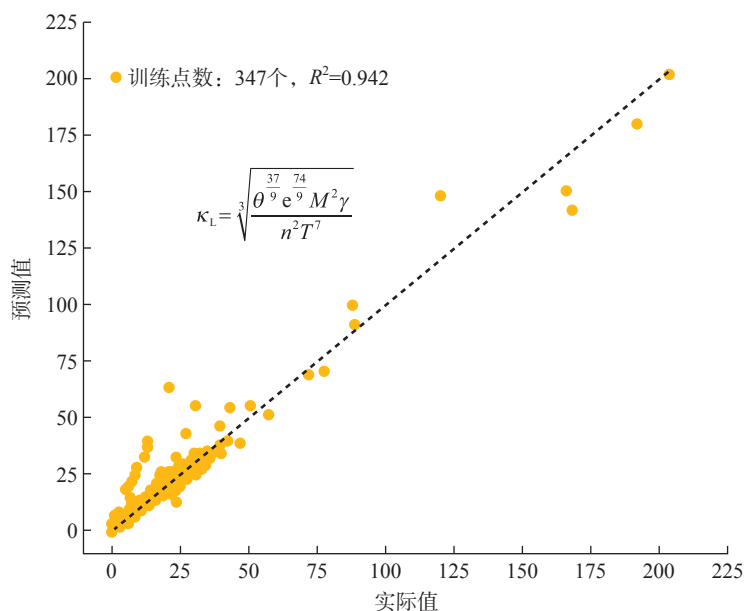


图 5 基于符号回归得到的计算公式与模型的预测结果

Fig.5 Calculation formula and results of model prediction based on symbolic regression

率是设计热障涂层的首要目标,目前机器学习的应用只集中于普通材料的晶格热导率,而忽略了涂层本身的微观结构和形态也是影响其导热能力的重要因素。

## 结论与展望

本文回顾了机器学习在航空发动机材料中的应用,主要关注了基于主动学习的材料开发和优化策略,对主动学习中每一个步骤进行了较为详细的总结,并系统回顾了机器学习或者数据驱动算法在高温合金、钛基合金、复合材料、热障涂层等方面的应用。虽然目前机器学习正在被越来越多地用于航空发动机材料研究中,但是依然存在大量的问题,需要用机器学习技术去解决。例如,目前的方法主要集中于利用机器学习构建成分、工艺与性能之间的预测模型,但是基于预测模型进行逆向设计的研究还比较鲜见。同时,航空发动机材料的组织结构对最终服役性能至关重要,目前的研究对组织结构的关注不够。将图像识别技术应用于材料微观组织结构的量化,并充分利用机器学习逆向设计的优势设计新材料是解决这些问题的关键。值得注意的是,因为航空发动机材料的服役环境恶劣,导致服役数据难以获得,即能够用于训练机器学习模型的数据集有限。这会导致数据模型的高度不确定性,如何解决数据的缺乏问题,是这一领域面临的关键问题。这一问题有望通过以下思路解决。

(1) 利用迁移学习,从容易获得的数据中挖掘信息,用于较少数据的建模,提升模型预测能力;

(2) 发展不确定性评估方法,对基于小数据的预测模型进行不确定性估计,有效利用不确定性来推荐候选材料或工艺,减少试验次数;

(3) 有效利用材料科学中存在的大量成熟理论如热力学和动力学

模型,提升数据模型性能,缩小待探索的未知材料/参数空间;

(4) 研发文本挖掘和自然语言处理技术,利用文献发表的数据,扩大样本数量。

总之,基于机器学习的材料研发是一种材料研究的新范式,将加速材料的开发效率,降低成本。这一新的研究方向结合了材料科学和数据科学、计算机科学、人工智能,是一个多学科交叉的研究领域,需要材料领域与计算机领域的专家相互协作,进一步推动其快速发展。

## 参考文献

[1] 刘大响. 一代新材料,一代新型发动机: 航空发动机的发展趋势及其对材料的需求[J]. 材料工程, 2017, 45(10): 1-5.

LIU Daxiang. One generation of new material, one generation of new type engine: Development trend of aero-engine and its requirements for materials[J]. Journal of Materials Engineering, 2017, 45(10): 1-5.

[2] 焦华宾, 莫松. 航空涡轮发动机现状及未来发展综述[J]. 航空制造技术, 2015, 58(12): 62-65.

JIAO Huabin, MO Song. Present status and development trend of aircraft turbine engine[J]. Aeronautical Manufacturing Technology, 2015, 58(12): 62-65.

[3] 王会阳, 安云岐, 李承宇, 等. 镍基高温合金材料的研究进展[J]. 材料导报, 2011, 25(S2): 482-486.

WANG Huiyang, AN Yunqi, LI Chengyu, et al. Research progress of Ni-based superalloys[J]. Materials Review, 2011, 25(S2): 482-486.

[4] 宿彦京, 付华栋, 白洋, 等. 中国材料基因工程研究进展[J]. 金属学报, 2020, 56(10): 1313-1323.

SU Yanjing, FU Huadong, BAI Yang, et al. Progress in materials genome engineering in China[J]. Acta Metallurgica Sinica, 2020, 56(10): 1313-1323.

[5] BUTLER K T, DAVIES D W, CARTWRIGHT H, et al. Machine learning for molecular and materials science[J]. Nature, 2018, 559(7715): 547-555.

[6] LIU Z K. First-principles calculations and CALPHAD modeling of thermodynamics[J]. Journal of Phase Equilibria and Diffusion, 2009,

30(5): 517-534.

[7] YU J X, GUO S, CHEN Y C, et al. A two-stage predicting model for  $\gamma'$  solvus temperature of L12-strengthened Co-base superalloys based on machine learning[J]. Intermetallics, 2019, 110: 106466.

[8] XUE D Z, XUE D Q, YUAN R H, et al. An informatics approach to transformation temperatures of NiTi-based shape memory alloys[J]. Acta Materialia, 2017, 125: 532-541.

[9] YUAN R H, LIU Z, BALACHANDRAN P V, et al. Accelerated discovery of large electrostrains in BaTiO<sub>3</sub>-based piezoelectrics using active learning[J]. Advanced Materials, 2018, 30(7): 1702884.

[10] LOOKMAN T, BALACHANDRAN P V, XUE D, et al. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design[J]. npj Computational Materials, 2019, 5(1): 227-235.

[11] AGGARWAL C C, ZHAI C X. Mining text data[M]. Boston: Springer Science & Business Media, 2012.

[12] GUPTA V, LEHAL G S. A survey of text mining techniques and applications[J]. Journal of Emerging Technologies in Web Intelligence, 2009, 1(1): 60-76.

[13] JAIN A, ONG S P, HAUTIER G, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation[J]. APL Materials, 2013, 1(1): 011002.

[14] CURTAROLO S, SETYAWAN W, HART G L W, et al. AFLOW: An automatic framework for high-throughput materials discovery[J]. Computational Materials Science, 2012, 58: 218-226.

[15] BELSKY A, HELLENBRANDT M, KAREN V L, et al. New developments in the inorganic crystal structure database (ICSD): Accessibility in support of materials research and design[J]. Acta Crystallographica Section B, 2002, 58(3): 364-369.

[16] PENG J, YAMAMOTO Y, HAWK J A, et al. Coupling physics in machine learning to predict properties of high-temperature alloys[J]. npj Computational Materials, 2020, 6(1): 141.

[17] RICKMAN J M, CHAN H M, HARMER M P, et al. Materials informatics for the screening of multi-principal elements and high-entropy alloys[J]. Nature Communications, 2019, 10(1): 316-323.

[18] KHATAVKAR N, SWETLANA S,

- SINGH A K. Accelerated prediction of Vickers hardness of Co- and Ni-based superalloys from microstructure and composition using advanced image processing techniques and machine learning[J]. *Acta Materialia*, 2020, 196: 295–303.
- [19] HU C Z, ZUO Y X, CHEN C, et al. Genetic algorithm-guided deep learning of grain boundary diagrams: Addressing the challenge of five degrees of freedom[J]. *Materials Today*, 2020, 38: 49–57.
- [20] SINGH D, SINGH B. Investigating the impact of data normalization on classification performance[J]. *Applied Soft Computing*, 2020, 97: 105524.
- [21] SEDGWICK P. Pearson's correlation coefficient[J]. *The BMJ*, 2012, 345: e4483. DOI:10.1136/bmj.e4483.
- [22] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. *Science*, 2011, 334(6062): 1518–1524.
- [23] MYERS L, SIROIS M J. Spearman correlation coefficients, differences between[J/OL]. *Encyclopedia of Statistical Sciences*, 2006. <https://doi.org/10.1002/0471667196.ess5050.pub2>
- [24] ABDI H. The Kendall rank correlation coefficient[J]. *Encyclopedia of Measurement and Statistics* Sage, 2007: 508–510.
- [25] LONGIN F, SOLNIK B. Extreme correlation of international equity markets[J]. *The Journal of Finance*, 2001, 56(2): 649–676.
- [26] LI R Z, ZHONG W, ZHU L P. Feature screening via distance correlation learning[J]. *Journal of the American Statistical Association*, 2012, 107(499): 1129–1139.
- [27] PARK T, CASELLA G. The Bayesian lasso[J]. *Journal of the American Statistical Association*, 2008, 103(482): 681–686.
- [28] DÍAZ-URIARTE R, ALVAREZ DE ANDRÉS S. Gene selection and classification of microarray data using random forest[J]. *BMC Bioinformatics*, 2006, 7(1): 3.
- [29] NOBLE W S. What is a support vector machine?[J]. *Nature Biotechnology*, 2006, 24(12): 1565–1567.
- [30] ENGEL E, DREIZLER R M. Relativistic density functional theory[M]. Berlin/Heidelberg: Springer-Verlag, 2013.
- [31] RUAN J J, XU W W, YANG T, et al. Accelerated design of novel W-free high-strength Co-base superalloys with extremely wide  $\gamma/\gamma'$  region by machine learning and CALPHAD methods[J]. *Acta Materialia*, 2020, 186: 425–433.
- [32] CHEN C T, MARTIN-MARTINEZ F J, LING S J, et al. Nacre-inspired design of graphene oxide-polydopamine nanocomposites for enhanced mechanical properties and multifunctionalities[J]. *Nano Futures*, 2017, 1(1): 011003.
- [33] XU W N, QIN Z, CHEN C T, et al. Ultrathin thermoresponsive self-folding 3D graphene[J]. *Science Advances*, 2017, 3(10): e1701084.
- [34] SHIN D, YAMAMOTO Y, BRADY M P, et al. Modern data analytics approach to predict creep of high-temperature alloys[J]. *Acta Materialia*, 2019, 168: 321–330.
- [35] YU J X, WANG C L, CHEN Y C, et al. Accelerated design of L12-strengthened Co-base superalloys based on machine learning of experimental data[J]. *Materials & Design*, 2020, 195: 108996.
- [36] ZOU M, LI W D, LI L F, et al. Machine learning assisted design approach for developing  $\gamma'$ -strengthened Co-Ni-base superalloys[M]. Cham: Springer International Publishing, 2020.
- [37] TAMURA R, OSADA T, MIN AGAWA K, et al. Machine learning-driven optimization in powder manufacturing of Ni-Co based superalloy[J]. *Materials & Design*, 2021, 198: 109290.
- [38] QIN Z J, WANG Z, WANG Y Q, et al. Phase prediction of Ni-base superalloys via high-throughput experiments and machine learning[J]. *Materials Research Letters*, 2021, 9(1): 32–40.
- [39] LUO J, LI M Q, YU W X. Prediction of flow stress in isothermal compression of Ti-6Al-4V alloy using fuzzy neural network[J]. *Materials & Design*, 2010, 31(6): 3078–3083.
- [40] SUN Z C, YANG H, TANG Z. Microstructural evolution model of TA15 titanium alloy based on BP neural network method and application in isothermal deformation[J]. *Computational Materials Science*, 2010, 50(2): 308–318.
- [41] LI Q, YU J Y, MU B C, et al. BP neural network prediction of the mechanical properties of porous NiTi shape memory alloy prepared by thermal explosion reaction[J]. *Materials Science and Engineering: A*, 2006, 419(1–2): 214–217.
- [42] SUN Y, ZENG W D, HAN Y F, et al. Determination of the influence of processing parameters on the mechanical properties of the Ti-6Al-4V alloy using an artificial neural network[J]. *Computational Materials Science*, 2012, 60: 239–244.
- [43] ARISOY Y M, ÖZEL T. Machine learning based predictive modeling of machining induced microhardness and grain size in Ti-6Al-4V alloy[J]. *Materials and Manufacturing Processes*, 2015, 30(4): 425–433.
- [44] HARSHA N, KUMAR I A, RAJU K S R, et al. Prediction of Machinability characteristics of Ti6Al4V alloy using Neural Networks and Neuro-Fuzzy techniques[J]. *Materials Today: Proceedings*, 2018, 5(2): 8454–8463.
- [45] ZACHARIA K, KRISHNAKUMAR P. Chatter prediction in high speed machining of titanium alloy (Ti-6Al-4V) using machine learning techniques[J]. *Materials Today: Proceedings*, 2020, 24: 350–358.
- [46] RAMKUMAR T, SELVAKUMAR M, CHANDRASEKAR P, et al. Monitoring the neural network modelling of wear behaviour of Ti-6Al-4V reinforced with nano  $B_4C$  particle[J]. *Materials Today: Proceedings*, 2021, 41: 942–950.
- [47] KHADEMI F, AKBARI M, JAMAL S M, et al. Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete[J]. *Frontiers of Structural and Civil Engineering*, 2017, 11(1): 90–99.
- [48] LI C L, NARAYANA P L, REDDY N S, et al. Modeling hot deformation behavior of low-cost Ti-2Al-9.2Mo-2Fe beta titanium alloy using a deep neural network[J]. *Journal of Materials Science & Technology*, 2019, 35(5): 907–916.
- [49] BHATTACHARYA S K, SAHARA R, NARUSHIMA T. Predicting the parabolic rate constants of high-temperature oxidation of Ti alloys using machine learning[J]. *Oxidation of Metals*, 2020, 94(3–4): 205–218.
- [50] ZHU C P, LI C, WU D, et al. A titanium alloys design method based on high-throughput experiments and machine learning[J]. *Journal of Materials Research and Technology*, 2021, 11: 2336–2353.
- [51] 付青峰, 杨细莲, 刘克明. 航空发动机高温材料的研究现状及展望[J]. *热处理技术与装备*, 2018, 39(3): 69–73.

FU Qingfeng, YANG Xilian, LIU Keming. Current status of research and prospect of high temperature materials for aeroengine[J]. Heat Treatment Technology and Equipment, 2018, 39(3): 69–73.

[52] PRAMOD R, VEERESH KUMAR G B, GOUDA P S S, et al. A study on the  $Al_2O_3$  reinforced Al7075 metal matrix composites wear behavior using artificial neural networks[J]. Materials Today: Proceedings, 2018, 5(5): 11376–11385.

[53] PATEL D K, PARTHASARATHY T, PRZYBYLA C. Predicting the effects of microstructure on matrix crack initiation in fiber reinforced ceramic matrix composites via machine learning[J]. Composite Structures, 2020, 236: 111702.

[54] CANAKCIA, VAROL T, OZSAHIN S. Artificial neural network to predict the effect of heat treatment, reinforcement size, and volume

fraction on AlCuMg alloy matrix composite properties fabricated by stir casting method[J]. The International Journal of Advanced Manufacturing Technology, 2015, 78(1–4): 305–317.

[55] CHEN L H, TRAN H, BATRA R, et al. Machine learning models for the lattice thermal conductivity prediction of inorganic materials[J]. Computational Materials Science, 2019, 170: 109155.

[56] JUNEJA R, YUMNAM G, SATSANGI S, et al. Coupling the high-throughput property map to machine learning for predicting lattice thermal conductivity[J]. Chemistry of Materials, 2019, 31(14): 5145–5151.

[57] LOFTIS C, YUAN K P, ZHAO Y, et al. Lattice thermal conductivity prediction using symbolic regression and machine learning[J]. The Journal of Physical Chemistry A, 2021, 125(1): 435–450.

[58] HALLOUA H, ELHASSNAOUI A, SAIFI A, et al. Neural networks and genetic algorithms for the evaluation of coatings thicknesses in thermal barriers by infrared thermography data[J]. Procedia Structural Integrity, 2017, 5: 997–1004.

[59] MA Z Y, ZHANG W, LUO Z B, et al. Ultrasonic characterization of thermal barrier coatings porosity through BP neural network optimizing Gaussian process regression algorithm[J]. Ultrasonics, 2020, 100: 105981.

[60] YE D D, WANG W Z, ZHOU H T, et al. Characterization of thermal barrier coatings microstructural features using terahertz spectroscopy[J]. Surface and Coatings Technology, 2020, 394: 125836.

通讯作者: 李金山, 教授, 博士, 研究方向为材料基因工程。

## Progress of Data-Driven Aero-Engine Materials Design

YUAN Ruihao, LIAO Weijie, TANG Bin, FAN Jiangkun, WANG Jun, KOU Hongchao, LI Jinshan

(State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an 710072, China)

**[ABSTRACT]** The data-driven materials design paradigm has been largely applied in the research of materials science, to shorten the process from materials discovery to applications. As the key and hot pot in the research of data-driven materials design, machine learning can extract the pattern or law behind the data, of which the active-learning based strategy has been successfully employed to guide the exploitation of materials. Firstly, this paper introduces the main idea and elements of the active learning loop, as well as the methods/algorithms used in each step. Then the paper reviews the applications and effect of machine learning in the field of aero-engine materials including superalloys, titanium-based alloys, composites and thermal barrier coating. Considering the extreme service environment of aero-engine materials, the review is concluded by discussing the possible challenges and solutions.

**Keywords:** Materials genome engineering; Machine learning; Active learning; Aero-engine materials; Superalloys

(责编 雷松)