

机器学习辅助的高通量实验加速硬质高熵合金 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 成分设计

王 炯¹, 肖 斌², 刘 轶^{1,2}

(1. 上海大学 材料基因组工程研究院, 上海 200444)

(2. 上海大学物理系 量子与分子结构国际中心, 上海 200444)

摘 要: 针对目标性能的多元合金成分设计因具有巨大的成分参数空间而极具挑战, 而且传统的试错实验由于效率低能探索的合金成分有限。提出利用高通量实验结合机器学习方法加速非等摩尔比的硬质高熵合金 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 的成分设计。首先通过自主研发的全流程高通量合金制备系统制备了 138 个不同成分的高熵合金铸态样品。然后根据测量的维氏硬度 (HV) 数据, 使用随机森林法和支撑向量机法进行机器学习建模, 并预测了五元合金体系内潜在的 3876 个不同成分合金的硬度。随机森林机器学习模型的预测结果在高 (HV>800 MPa)、中 (600<HV<800 MPa)、低 (HV<600 MPa) 硬度区域的平均误差分别为 2.87%, 3.30% 和 6.70%, 实验硬度值在对应区域的测量误差分别为 1.69%, 1.88% 和 1.87%。根据机器学习模型预测结果建立的“成分-硬度”与“描述因子-硬度”关系图谱展示了全成分空间内高熵合金的硬度变化规律及影响硬度的重要描述因子——原子半径差。研究结果表明, 高通量实验与机器学习相结合可使多元合金成分优化效率提高百倍以上。此外, 建议未来研究应在“机器学习”基础上加强“向机器学习”, 在更高层次上获得新的专业知识认知。

关键词: 高通量实验; 机器学习; 高熵合金; 硬度

中图分类号: TP181; TG146 **文献标识码:** A **文章编号:** 1674-3962(2020)04-0269-09

Machine Learning Assisted High-Throughput Experiments Accelerates the Composition Design of Hard High-Entropy Alloy $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$

WANG Jiong¹, XIAO Bin², LIU Yi^{1,2}

(1. Materials Genome Institute, Shanghai University, Shanghai 200444, China)

(2. International Centre for Quantum and Molecular Structures, Department of Physics, Shanghai University, Shanghai 200444, China)

Abstract: The composition design of multi-component alloy for the target performance is extremely challenging due to the enormous potential composition. The traditional trial-and-error experiments can only explore limited alloy compositions because of its low efficiency. In this work, the composition design of non-equimolar hard high-entropy alloy $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ was accelerated via combining the high-throughput experiment with machine learning. Firstly, 138 as-cast high-entropy alloys were prepared by a home-developed all-process high-throughput alloy synthesis system. Then, the machine learning models were built based on the measured Vickers hardness (HV) by using random forest (RF) and supporting vector machine methods. And, they made the prediction of HV values for 3876 potential alloys in the five-component alloy system. The HV values predicted by RF machine learning models have the averaged errors of 2.87%, 3.30% and 6.70%, respectively in high (HV>800 MPa), medium (600<HV<800 MPa), and low (HV<600 MPa) hardness regions. The measurement errors of experimental data at the corresponding hardness regions were 1.69%, 1.88% and 1.87%, respectively. The “composition-hardness” and “descriptor-hardness” relationship maps, constructed by using the prediction of machine learning models, show the variation tendency of alloy hardness at the full composition space and the important descriptors to alloy hardness, namely atom size difference. This work demonstrates that the high-throughput experiments combined with machine learning can accelerate the composition design of multi-component alloy more than hundred times. Additionally, future research should emphasize on “learning from machine” based on “machine

收稿日期: 2019-05-26

基金项目: 国家科技部重点研发计划“材料基因组工程”项目
(2017YFB0702901, 2017YFB0701502); 国家自然科学基金项目 (91641128)

第一作者: 王 炯, 男, 1990 年生, 硕士研究生

通讯作者: 刘 轶, 男, 1971 年生, 教授, 博士生导师,

Email: yiliu@t.shu.edu.cn

DOI: 10.7502/j.issn.1674-3962.201905032

of experimental data at the corresponding hardness regions were 1.69%, 1.88% and 1.87%, respectively. The “composition-hardness” and “descriptor-hardness” relationship maps, constructed by using the prediction of machine learning models, show the variation tendency of alloy hardness at the full composition space and the important descriptors to alloy hardness, namely atom size difference. This work demonstrates that the high-throughput experiments combined with machine learning can accelerate the composition design of multi-component alloy more than hundred times. Additionally, future research should emphasize on “learning from machine” based on “machine

learning” to acquire the new domain knowledge at a higher level.

Key words: high-throughput experiment; machine learning; high-entropy alloy; hardness

1 前言

研发具有特定目标性能的新材料时,传统的材料研究通常采用经验试错法,通过少量实验离散地调整样品成分与制备工艺来探索材料的最佳性能。然而,这种“炒菜式”研发手段耗时且成本高昂,同时由于实验数据量较少,对材料结构-性能关系考察不够全面深入。随着材料基因组工程的兴起^[1],利用高通量实验、计算模拟和机器学习加速新材料研发,受到了越来越多的关注。早在 1970 年, Hanak^[2] 提出了包含高通量材料研发 (high-throughput materials development, HTMD) 在内的所有环节流程,该研究被认为是材料领域内最早提出高通量概念的工作。1980 年, Moulijn 等^[3] 发表了使用多相催化平行反应器的高通量材料制备研究。1997 年, Danielson 等^[4] 采用电子束蒸发法在一个直径为 7.6 cm 的衬底上合成了 25 000 种不同成分的材料。近年来,研究者们基于材料组合芯片思想,采用气相沉积、磁控溅射等工艺制备了每个成分连续变化的多组元薄膜样品或一批成分离散变化的块体样品^[5-9]。总之,高通量材料制备方法具有多工位、自动化、并发式的特点,可以实现材料的快速、批量合成,比传统的单一样品生产模式在材料制备效率上有大幅度提升。随着实验数据的积累,材料数据库在新材料研发设计过程中也起到越来越重要的作用。Bligaard 等^[10] 利用包含晶格尺寸、体积模量等物理量的有序化金属玻璃数据库,筛选出了具有低压缩性、高稳定性和低成本的备选材料。但通常情况下,针对某一类材料的单一或几个性能进行设计时,高质量、大批量的数据还很缺乏,甚至某些全新体系可能不存在已有实验数据,此时高通量实验成为获取大量优质、自治实验数据的有力手段。

虽然高通量实验能够有效提升研究效率,但面对多元合金巨大的成分参数空间(数千至数万个成分组合)时,探索所有潜在合金成分在实验上是几乎不可能的,合金成分优化设计依然是一项艰巨的挑战。因此,将机器学习法引入到材料成分设计领域成为了加速新材料研发进程的重要手段。传统的机器学习法中,以神经网络(ANN)、支持向量机(SVM)、随机森林(RF)、K 近邻(KNN)等方法最具代表性^[11-22]。2016 年, Raccuglia 等^[23] 利用失败实验数据结合 SVM 算法成功建立了无机-有机杂化材料结晶过程反应模型,并预测了新化合物的形成条件,准确率达 89%。此外,机器学习法也被作为加速研究方法广泛应用于介电特性^[24]、无机晶体带

隙^[25]、熔点^[26]、抗压强度^[27]、硬度^[28]、相结构^[29]等材料性质研究。

自 Senkov 等^[30, 31] 制备了具有优异高温性能的 W-Nb-Mo-Ta-V 系列 bcc 相难熔高熵合金后,关于难熔高熵合金的研究受到了极大关注。研究者们广泛研究了含 W, Mo, Ti, Nb, Ta, Zr 的高熵合金体系^[32-37]。除纯相高熵合金外,还广泛研究了高熵合金中第二相对合金性能的影响^[38-40]。本工作将高通量实验与机器学习法相结合,应用于开发新型非等摩尔比的硬质高熵合金。在元素组成上,考虑到高熔点的 W 和 Mo 以及 Co, Cr, Ti 等在高温结构材料中广泛应用的金属元素,以 CoCrTiMoW 体系为基础设计了新型非等摩尔比的硬质高熵合金。

2 实验

2.1 高通量实验

本工作首次提出了以制备离散成分块体合金为特色的全流程高通量材料制备概念,涵盖合金样品配料、混合、压块、电弧熔炼、镶嵌、切割、磨抛的合金熔炼和金相样品制备环节。基于此思想,上海大学材料基因组工程研究院与合肥/沈阳科晶材料技术有限公司联合设计研发了全流程高通量合金制备系统,具体包括 36 工位粉料分配器、16 工位球磨混料机、16 工位压力机、32 工位自动电弧熔炼炉、16 工位冷镶嵌机、8 工位线切割机以及 16 工位自动磨抛机(图 1)。本实验的合金熔炼和样品制备均由高通量合金制备系统完成,对材料制备流程各环节实现了多工位、自动化加速,缩减了单个样品的平均实验准备和运行时间,例如省略了传统电弧熔炼的多次抽真空过程,克服了材料制备过程中耗时的瓶颈。该高通量合金制备系统的整体效率比传统单一或少样品制备方法提升了至少 10 倍。

本工作以粒径均约为 50 μm 的 Co, Cr, Ti, Mo, W 纯金属粉为原材料,其纯度均大于 99.5% (质量分数),共制备了 138 个合金样品,其成分分布在 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 空间内。初始的合金成分设计采用以下两种策略:① $\text{CoCrTiMo}_u\text{W}_v$, 重点调整难熔金属 W 和 Mo 的含量;② $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$, 首先将成分映射到与合金相稳定性有关的描述因子价电子浓度-原子半径差 (VEC- δ) 空间^[41],然后选取 VEC- δ 空间内均匀分布的点转换回成分。

本工作使用自动粉料分配器根据给定合金成分进行配粉。将配置好的合金粉末经过自动球磨机充分混合均匀后,利用自动压力机压制成圆柱状块体(直径约为 2.05 cm,高度约为 1.50 cm)。然后在 Ar 气氛保护下的

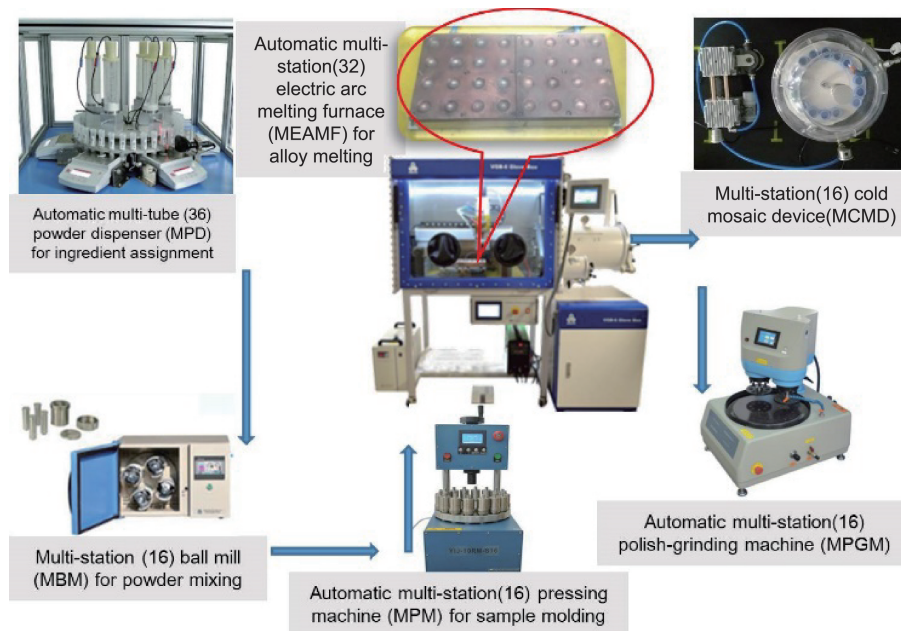
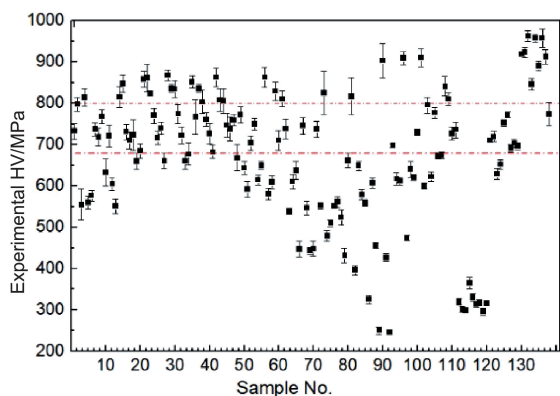


图1 全流程高通量合金制备系统

Fig. 1 All-process high-throughput synthesis system of alloy

手套箱内利用自动电弧熔炼炉中的水冷铜模具自动连续熔炼制备铸态金属锭。每个工作流程最多可熔炼 32 个金属锭, 单个金属锭的重量约为 10 g, 需反复熔炼 5 次以上以保证其组织均匀性。然后, 经冷镶嵌和磨抛后(依次用颗粒粒度为 70, 37, 15, 12, 7 和 5 μm 的砂纸打磨、抛光)制成金相样品。最后, 采用 HBRV(D)-187.5A1 型手动布洛维三用硬度计测试合金的宏观维氏硬度(HV), 测试条件为: 压力为 30 kg, 预加载时间为 3 s, 加载时间为 10 s。对每个样品进行多次宏观硬度测量, 至少获得五次以上有效值并取其平均值。138 个不同成分的合金的 HV 测量结果如图 2 所示, 其中 36 个合金的 HV 低于 600 MPa, 70 个合金的 HV 在 600~800 MPa 之间, 32 个合金的 HV 高于 800 MPa。

图2 138 个高熵合金 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 样品的硬度实验测量值Fig. 2 Experimental HV values of 138 high-entropy alloy $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ samples

2.2 机器学习

机器学习数据集为通过高通量实验获得的 138 个合金的硬度数据。由于全流程高通量合金制备系统尽可能提供了相同的实验环境和过程, 减少了因工艺参数不同造成的数据不确定性。对于不同的机器学习预测模型, 本工作采用决定系数 R^2 、平均绝对误差 MAE 与均方根误差 RMSE 来综合评判模型优劣。考虑到小数据样本和数据本身的误差, 在最终模型选用上, 采用评判拟合效果的相关系数 R^2 作为第一判据, 并结合考虑 MAE 和 RMSE。 R^2 , MAE 和 RMSE 由式(1)~式(3)计算得到:

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

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

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

其中, y_i 是数据集中 HV 的实验测量值, \bar{y} 是 y_i 的平均值, \hat{y}_i 为 HV 的模型预测值。

2.2.1 机器学习算法与特征

本工作测试了 RF^[42] 和 SVM^[43] 机器学习算法, 其中 SVM 法选择了高斯径向基核函数(rbf)与线性核函数(linear), 将上述算法分别记为 RF、SVM_rbf 和 SVM_linear。所有机器学习算法由 Python 语言实现, 主要使用 Scikit-learn 算法包^[44] 以及 Pandas 和 Numpy 计算扩展库实现数据的读入与处理。

本工作的机器学习共使用了 29 个基础特征描述因子(表 1)^[45-48]和 1 个目标变量——HV。机器学习模型的输入特征是基于合金体系中不同元素的基本属性,以各组分成分的摩尔比为权重的算术平均值构建生成。因此,将 29 个基础元素或单质体相材料的描述因子分为 4 组:

(1)AD: 编号 1~29 的所有描述因子。

(2)EMF: 只包含编号 1~5 表示元素摩尔比的描述因子。

(3)VDSHOT: VEC, ΔH_{mix} , ΔS_{mix} , T_{m} , Ω , δ 与高熵合金相结构关联的特征^[41, 49-52]。

(4)SD: 选择特征集合,由经 RF 算法评估后排名前 9 位的描述因子组成(δ , DC, NC, RC, ΔS_{mix} , EI, W, EN, Cr)。

表 1 机器学习模型的描述因子^[45-48]
Table 1 Descriptors of machine learning models^[45-48]

| No. | Symbol | Definition | No. | Symbol | Definition |
|-----|-------------------------|---|-----|--------|--|
| 1 | Co | Cobalt mole fraction | 16 | DC | Distance, core electron (Schubert) ^[48] |
| 2 | Cr | Chromium mole fraction | 17 | DV | Distance, valence electron (Schubert) moment ^[48] |
| 3 | Ti | Titanium mole fraction | 18 | EN | Electronegativity (Martynov and Batsanov) ^[48] |
| 4 | Mo | Molybdenum mole fraction | 19 | EC | Cohesive energy (Brewer) ^[48] |
| 5 | W | Tungsten mole fraction | 20 | EI | Energy of first ionization ^[48] |
| 6 | VEC | Valence electron concentration ^[45] | 21 | EM | Enthalpy of melting ^[48] |
| 7 | ΔH_{mix} | Enthalpy of mixing | 22 | MP | Mendeleev number (Pettifor) ^[48] |
| 8 | ΔS_{mix} | Entropy of mixing | 23 | NC | Effective nuclear charge ^[48] |
| 9 | T_{m} | Average melting point | 24 | QN | Quantum number ^[48] |
| 10 | Ω | $\Omega = (T_{\text{m}} \cdot \Delta S_{\text{mix}}) / \Delta H_{\text{mix}} $ ^[46] | 25 | RC | Covalent radius ^[48] |
| 11 | δ | Atomic size difference ^[47] | 26 | RM | Metal radius (Waber) ^[48] |
| 12 | AN | Atomic number ^[48] | 27 | TM | Average melting point ^[48] |
| 13 | AW | Atomic weight ^[48] | 28 | TN | Thermal neutron capture cross section ^[48] |
| 14 | CH | Charge, nuclear effective charge(Clementi) ^[48] | 29 | BM | Bulk modulus ^[48] |
| 15 | DS | Density ^[48] | | | |

2. 2. 2 机器学习策略

将 3 种机器学习算法与 4 种描述因子集合进行组合,构建了 12 个机器学习方法,如表 2 所示。使用五折交叉验证法分别对 12 个机器学习方法的超参数进行优化,即 80%的数据用于建模,20%的数据用于评估算法的精度。对 12 个机器学习方法分别进行 10 次随机数据集划分(8:2)建模,共获得 120 个机器学习模型,记为 ML(120)。

表 2 由 3 种机器学习算法与 4 种描述因子集合组合而成的 12 个机器学习方法

Table 2 Twelve machine learning methods combined by four descriptor sets and three algorithms

| Descriptor set | Algorithm | | |
|----------------|----------------|-------------------|-----------|
| | SVM_rbf | SVM_linear | RF |
| AD | SVM_rbf/AD | SVM_linear/AD | RF/AD |
| EMF | SVM_rbf/EMF | SVM_linear/EMF | RF/EMF |
| SD | SVM_rbf/SD | SVM_linear/SD | RF/SD |
| VDSHOT | SVM_rbf/VDSHOT | SVM_linear/VDSHOT | RF/VDSHOT |

2. 2. 3 模型选择

图 3 为按照描述因子集合划分的 120 个机器学习模型

评估结果。RF、SVM_rbf 和 SVM_linear 3 种机器学习算法使用 4 种描述因子集合构建的模型的预测能力定性基本相似,其中以 AD 构建的模型的预测结果定量最优,后续针对该系列模型的预测结果进一步讨论。不同机器学习算法的预测精度排序为 RF>SVM_rbf>SVM_linear。虽然 RF 算法的模型预测能力最好,但 SVM 算法在随数据集划分变化时的波动性显著小于 RF 算法,这说明 SVM 算法相较于 RF 算法对小样本数据集具有更强的鲁棒性。考虑到机器学习建模过程中小样本数据集划分对模型的影响,在最终模型的选择上除了以 R²作为判据,也结合较低的 MAE 和 RMSE 进行综合评判,因此挑选出的最佳模型为 RF/AD_4。此外,还从鲁棒性强的 SVM_rbf 算法生成的模型中依据 R²选出一个最优模型 SVM_rbf/AD-5,进而比较这两种模型在合金硬度预测趋势上的异同。

预测精度最好的 RF/AD_4 和 SVM_rbf/AD-5 模型的评估结果如表 3 所示,138 个合金样品硬度的实验测量值(H138e)和机器学习模型预测值(H138m)如图 4 所示。两个模型的预测结果在趋势上表现出一致性,但 RF/AD_4

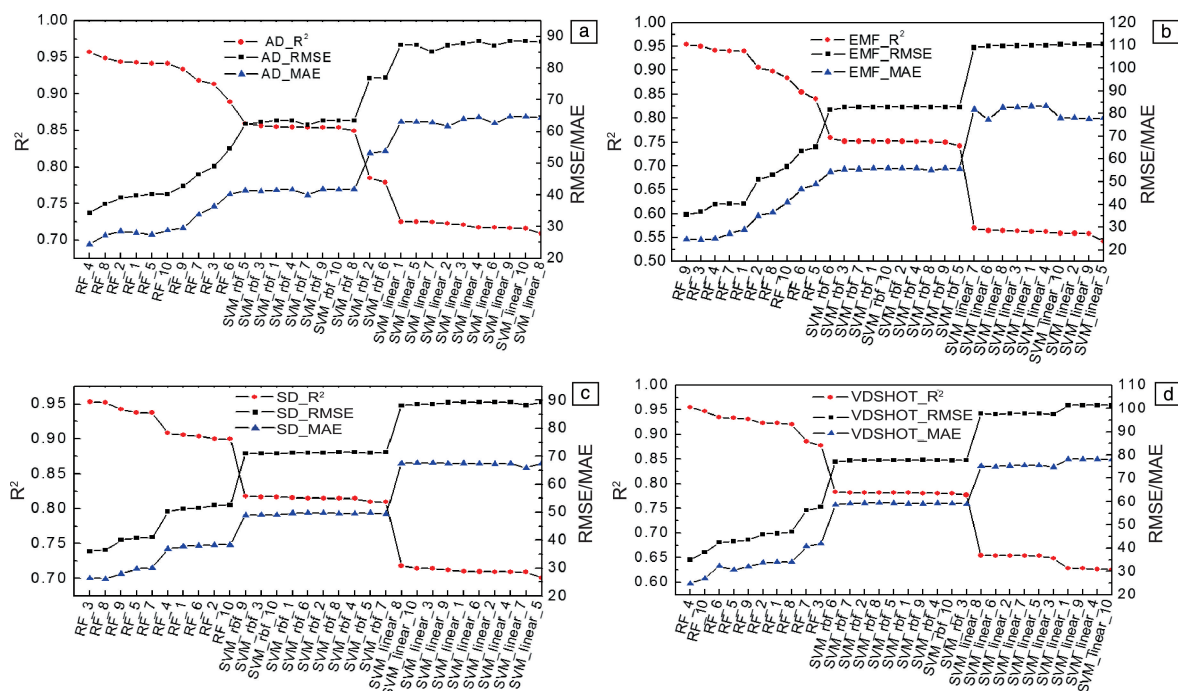


图3 按照描述因子集合划分的120个机器学习模型的评估结果：(a)AD, (b)EMF, (c)SD, (d)VDSHOT

Fig. 3 The evaluation results of 120 machine learning models divided by different descriptor sets: (a) AD, (b) EMF, (c) SD, (d) VDSHOT

模型在数据相关性和预测误差方面明显要优于SVM_rbf/AD-5模型。RF/AD_4模型在低($\text{HV} < 600 \text{ MPa}$)、中($600 < \text{HV} < 800 \text{ MPa}$)和高($\text{HV} > 800 \text{ MPa}$)硬度区间的相对误差分别为6.70%、3.30%和2.87%，而实验测量值的相对误差在对应区域分别为1.87%、1.88%和1.69%。在中、高硬度区，RF/AD_4模型的硬度预测值和实验测量值比较接近，但其在低硬度区域的预测准确性相对较低。这可能是因为在低硬度区域的数据量较少，且该区域对应的合金成分多以单组元为主，可能导致合金结构向以该单组元为基体的单相固溶体转变，从而使该合金在某一成分临界值处出现性能突变。因此，为了进一步提高机器学习的精度，需要增加误差较大部分的实验数

据。另外，机器学习模型的预测值和实验平均值间的差距，与实验多点测量值同其平均值的差距非常接近，表明模型具有预测其他未测量合金硬度的能力，而且还可以系统了解 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 合金体系中成分对其硬度的影响规律。

表3 RF/AD_4与SVM_rbf/AD-5模型的评估结果

Table 3 The evaluation results of RF/AD_4 and SVM_rbf/AD-5 models

| Model | R^2 | MAE | RMSE | Relative error | | |
|--------------|-------|-------|-------|----------------|--------------------|--------------|
| | | | | HV < 600 MPa | 600 < HV < 800 MPa | HV > 800 MPa |
| RF/AD_4 | 0.96 | 24.41 | 34.33 | 6.70% | 3.30% | 2.87% |
| SVM_rbf/AD-5 | 0.86 | 41.38 | 62.38 | 8.30% | 6.30% | 5.15% |

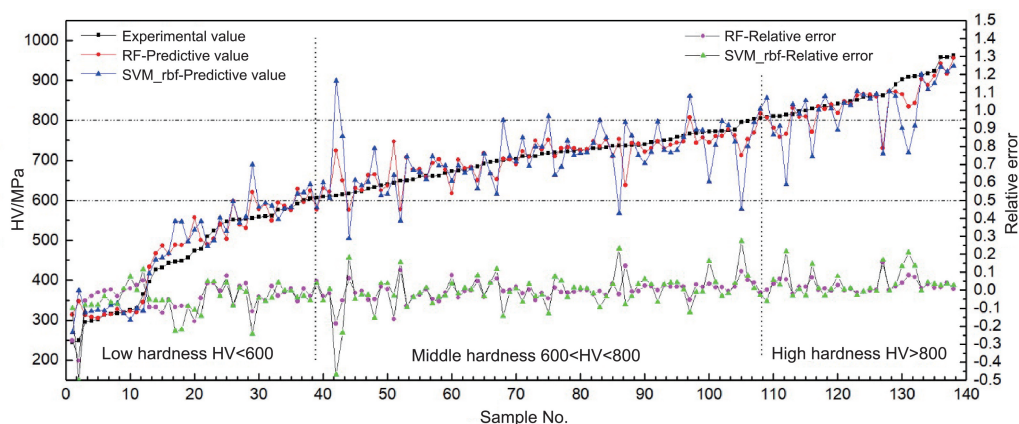


图4 138个合金样品硬度的实验测量值、模型预测值以及预测值的相对误差

Fig. 4 Experimental values, predictive values and relative errors of predictive values to the hardness of 138 alloy samples

3 结果分析

3.1 描述因子的选择

使用 RF 算法对 29 个描述因子的重要性进行排序 (图 5), 其中 RF 算法的回归树采用最小均方差原则计算、评估描述因子的重要性, 其重要性由高到低排名前 9 的分别为: δ , DC, NC, RC, ΔS_{mix} , EI, W, EN 和 Cr。根据重要性排序, 其中最重要的描述因子为 δ , 表明 δ 在 29 个描述因子中对合金硬度的影响最大, 相关文献也证明了 δ 对合金硬度有显著影响^[53, 54]。从合金组元角度分析, 各元素按其对合金硬度影响的大小排序为: W>Cr>Mo>Ti>Co。因此, 将排名前 9 的描述因子作为 SD 用于机器学习建模, 由其构建的模型预测结果如图 3c 所示。

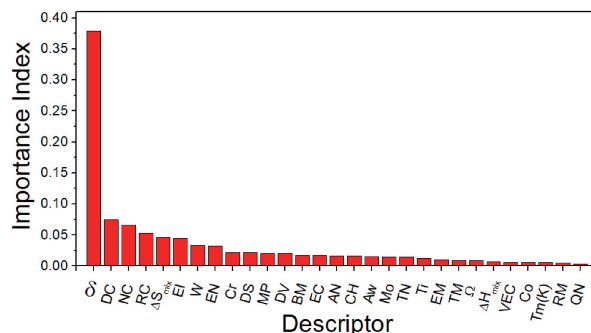


图 5 机器学习中 29 个描述因子的重要性排序

Fig. 5 Importance order of 29 descriptors in machine learning

3.2 “成分-硬度”图谱

图 6 分别为根据 RF/AD_4 与 SVM_rbf/AD-5 模型预测结果绘制的高熵合金全成分空间的“成分-硬度”图谱

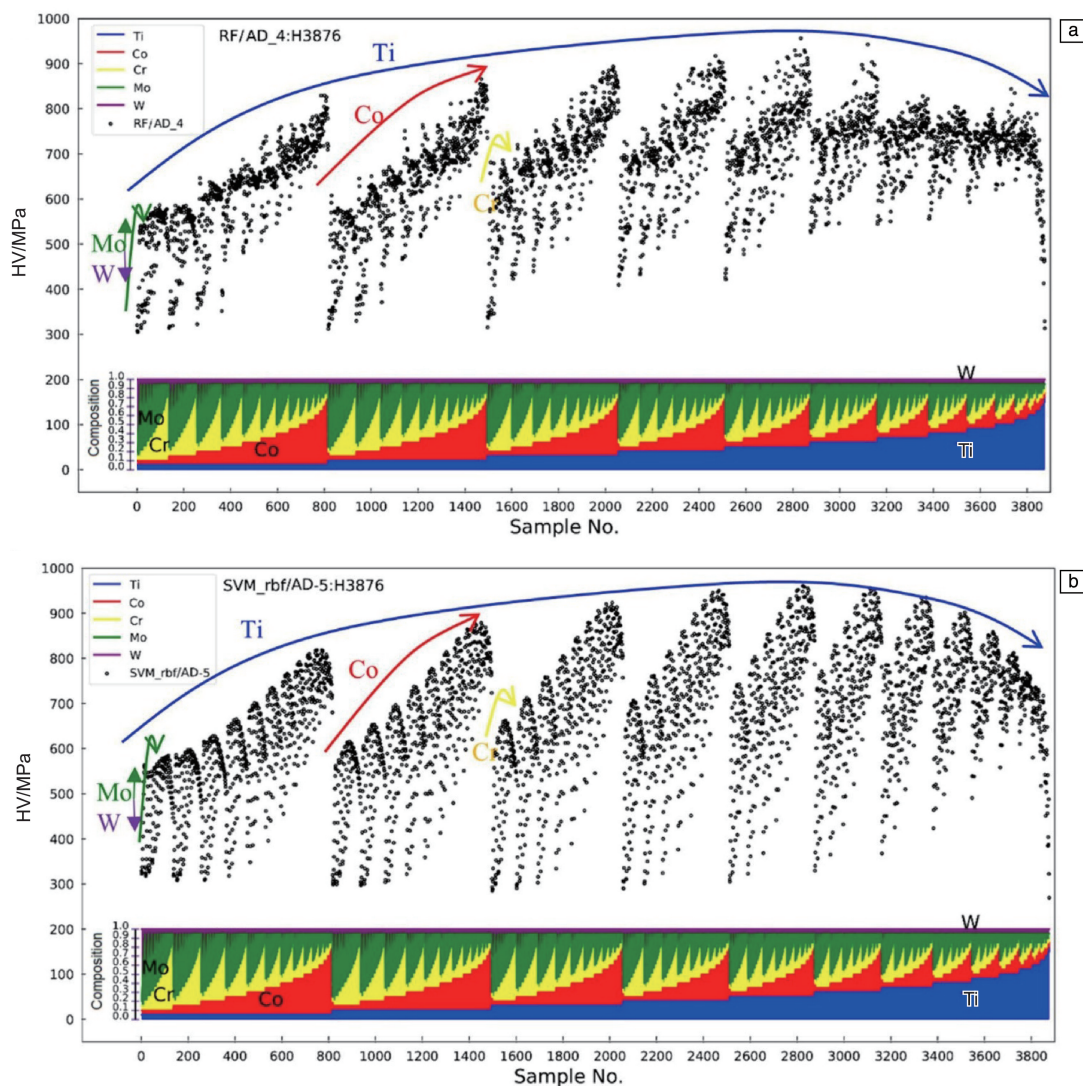


图 6 根据两种机器学习模型预测结果绘制的高熵合金“成分-硬度”图谱: (a) RF/AD_4, (b) SVM_rbf/AD-5

Fig. 6 “Composition-hardness” maps of high-entropy alloy constructed by the two kinds of machine learning models: (a) RF/AD_4, (b) SVM_rbf/AD-5

(composition-hardnessmap, CH map)。两个模型在对 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 合金体系全成分空间的硬度预测结果上显示出相似的趋势。对于五组元成分体系, 有 4 个独立成分变量。两个模型在 CH 图谱上都具有 4 条与成分变量相对应的硬度包络曲线, 且各硬度峰值对应出现在相同的成分区间内。以合金元素的摩尔比作为变量, 硬度随成分改变呈现有规律的变化趋势。

以 Ti 作为第一层变化元素, 可以看到合金体系的硬度包络曲线随着 Ti 含量的升高先升高再下降, 而且后半段曲线有加速下降的趋势。结合一般二元固溶体合金的经验, 在一种元素的含量由 0 增加到 1 的过程中, 合金一般会经历“单相-固溶相-金属间相-固溶相-单相”的转变过程。在单相区间内, 随着元素含量的变化, 晶格畸变也将发生显著变化, 从而影响位错在晶粒内部的运动, 产生位错钉扎等效应, 对合金的力学性能产生影响。当

合金内产生金属间相, 或单相结构产生较大的晶格畸变时, 合金会表现出较高的硬度; 而当合金成分接近单一组元(即固溶组元含量少)、晶格畸变较小时, 合金一般表现出较低的硬度。由图 6 中的 4 条硬度包络曲线可以看出, 在某一元素的含量达到最高值时, 其对应的合金硬度出现跳变, 由于 x 轴的成分变量按照 Ti, Co, Cr, Mo, W 的顺序进行排序组合, 因此 Ti, Co, Cr 对应的 3 条硬度包络曲线变化更为显著。这两种模型的预测结果与传统经验有类似的变化规律, 这也证明了模型预测结果的可靠性。

3.3 “描述因子-硬度”图谱

除了建立 CH 图谱, 还根据 RF/AD_4 模型预测结果, 将硬度投影到二维机器学习描述因子平面中构建“描述因子-硬度”图谱(descriptor-hardness map, DH map), 如图 7 所示。由于部分描述因子具有特定的物理含义,

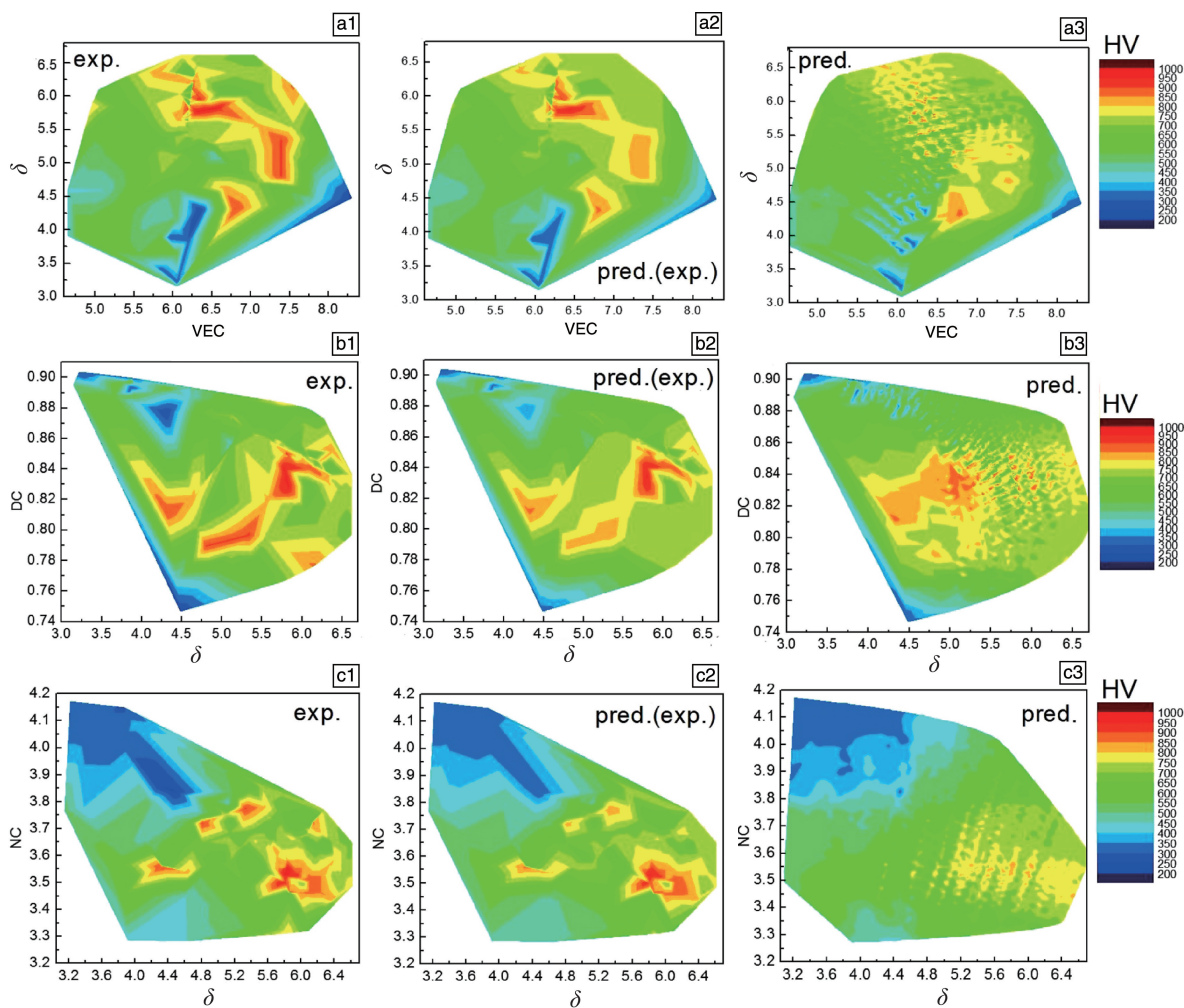


图 7 138 个合金硬度的实验测量值(H138e)(1)、机器学习模型预测(H138m)(2)和全局成分空间内 3876 个合金硬度的模型预测值(H3876m)(3)的“描述因子-硬度”图谱(图中的彩色标尺代表硬度, 从蓝到红表示硬度增加): (a) VEG- δ , (b) δ -DC 和 (c) δ -NC

Fig. 7 “Descriptor-Hardness” maps of H138e (1), H138m (2) and H3876m (3) (The color scales represent the hardness value increased from blue to red): (a) VEG- δ , (b) δ -DC and (c) δ -NC

这种表示材料性质的描述因子投影图有助于理解合金成分之外的描述因子对合金硬度的影响规律。这种不含特定合金成分的图谱可能具有更广泛的适用性, 不仅可应用于 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 特定合金体系, 还可能扩展到其他多元合金体系的硬度预测。另外, 将多组元成分空间投影到二维描述因子空间内, 还将有效降低合金初始成分选择的维度, 实现快速简便的合金成分反向设计。

图 7a1~7c1 为 H138e 分别在 VEG- δ 、 δ -DC 和 δ -NC 空间的 DH 图谱; 图 7a2~7c2 为 H138m 分别在 VEG- δ 、 δ -DC 和 δ -NC 空间的 DH 图谱; 图 7a3~7c3 为全成分空间内 3876 个合金硬度的模型预测值 (H3876m) 分别在 VEG- δ 、 δ -DC 和 δ -NC 空间的 DH 图谱。与 H3876m 的 DH 图谱相比, H138m 与 H138e 的 DH 图谱更吻合, 这是因为后两者是点对点对应, 而前者在相同范围内布点更密集, 其高硬度与低硬度可能非常接近。尽管预测结果在绝对数值上可能存在差异, 但在二维描述因子空间下, 硬度分布呈现出定性相似模式。在 DH 图谱中, VEG- δ 映射中的高硬度区和低硬度区间的分离距离最大, 表明 VEG- δ 空间在分离和预测硬度上具有优势。实际硬度预测可以由多个 DH 映射同时确定, 进一步增加预测的可靠性。

4 结 论

本文利用高通量实验结合机器学习的方法加速非等摩尔比的硬质高熵合金 $\text{Co}_x\text{Cr}_y\text{Ti}_z\text{Mo}_u\text{W}_v$ 的成分优化设计, 使综合设计效率提高了 200 倍以上。利用本团队自主研发的一系列新型全流程高通量合金制备实验设备, 多工位、大批量、自动化制备具有离散成分的块体合金, 较传统单/少样品制备过程至少加速 10 倍。该套高通量材料制备系统包括合金熔炼和金相样品制备流程, 具体涵盖配料、混合、压块、电弧熔炼、镶嵌、切割和磨抛。利用高通量实验合成了 138 个不同成分的合金样品, 根据其硬度数据使用 3 种机器学习算法 (随机森林和两种支持向量机法) 和 4 种描述因子集合构建了 120 个机器学习模型。研究表明, 随机森林比支持向量机精度更高, 其预测结果与实验测量值接近。通过机器学习建模设计合金成分, 与在全成分空间内排列组合式的穷尽搜索相比至少加速了 20 倍。利用机器学习模型构建全成分空间内的“成分-性质”和“描述因子-性质”关系, “从数据驱动设计再返回到知识驱动设计”, 从而实现认知的螺旋式上升。本工作证明了材料基因组计划提出的低成本加倍材料研发速度不仅可能, 而且效率会比最初提议的两倍更高。机器学习指导下的高通量实验方法可成为加速多组元材料成分优化设计的有效通用策略。未来材料研究需

要在“机器学习”的基础上聚焦“向机器学习”, 从而获得新的专业领域知识。

致谢: 感谢上海大学高水平大学建设项目对高通量实验设备研制的支持; 感谢中铝材料应用研究院、鞍山钢铁公司、云南锡业集团、福建南平铝业公司和上海紫燕合金公司对高通量材料研发的支持; 同时感谢科晶 (MTI) 集团的江晓平博士在共同研发高通量合金制备系统方面的帮助。

参考文献 References

- [1] 王卓, 王礞, 雍歧龙, 等. 中国材料进展[J], 2017, 36(2): 132-140.
WANG Z, WANG M, YONG Q L, *et al.* Materials China[J], 2017, 36(2): 132-140.
- [2] HANAK J J. Journal of Materials Science[J], 1970, 5(11): 964-971.
- [3] THOMAS R, MOULIJN J A, DE BEER V H J, *et al.* Journal of Molecular Catalysis[J], 1980, 8(1-3): 161-174.
- [4] DANIELSON E, GOLDEN J H, MCFARLAND E W, *et al.* Nature[J], 1997, 389(6654): 944-948.
- [5] BRICEÑO G, CHANG H Y, SUN X D, *et al.* Science[J], 1995, 270(5234): 273-275.
- [6] WANG J S, YOO Y, GAO C, *et al.* Science[J], 1998, 279(5357): 1712-1714.
- [7] XIANG X D, SUN X D, BRICEÑO G, *et al.* Science[J], 1995, 268(5218): 1738-1740.
- [8] XIANG X D, WANG G, ZHANG X K, *et al.* Engineering[J], 2015, 1(2): 225-233.
- [9] LIU Y H, HU Z H, SUO Z G, *et al.* Science China Technological Sciences[J], 2019, 62(4): 521-545.
- [10] BLIGAARD T, JOHANNESSON G H, RUBAN A V, *et al.* Applied Physics Letters[J], 2003, 83(22): 4527-4529.
- [11] DE JONG M, CHEN W, GEERLINGS H, *et al.* Scientific Data[J], 2015, 2: 150053.
- [12] ONG S P, WANG L, KANG B, *et al.* Chemistry of Materials[J], 2008, 20(5): 1798-1807.
- [13] ONG S P, RICHARDS W D, JAIN A, *et al.* Computational Materials Science[J], 2013, 68: 314-319.
- [14] BRUN F, YOSHIDA T, ROBSON J D, *et al.* Materials Science and Technology[J], 1999, 15(5): 547-554.
- [15] SOURMAIL T, BHADSHIA H K D H, MACKAY D J C. Materials Science and Technology[J], 2002, 18(6): 655-663.
- [16] JOO M S, RYU J H, BHADSHIA H K D H. Materials and Manufacturing Processes[J], 2008, 24(1): 53-58.
- [17] KHALAJ G, NAZARI A, POURALIAKBAR H. Neural Network World[J], 2013, 23(2): 117-130.
- [18] KHALAJ G. Neural Computing and Applications[J], 2013, 23(3/4): 779-786.

- [19] FAIZABADI M J, KHALAJ G, POURALIAKBAR H, *et al.* Neural Computing and Applications[J], 2014, 25(7/8): 1993–1999.
- [20] POURALIAKBAR H, KHALAJ M J, NAZERFAKHARI M, *et al.* Journal of Iron and Steel Research International[J], 2015, 22(5): 446–450.
- [21] CONDUIT B D, JONES N G, STONE H J, *et al.* Materials & Design[J], 2017, 131: 358–365.
- [22] ISLAM N, HUANG W J, ZHUANG H L L. Computational Materials Science[J], 2018, 150: 230–235.
- [23] RACCUGLIA P, ELBERT K C, ADLER P D F, *et al.* Nature[J], 2016, 533(7601): 73–77.
- [24] PILANIA G, WANG C C, JIANG X, *et al.* Scientific Reports[J], 2013, 3: 2810.
- [25] ISAYEV O, OSES C, TOHER C, *et al.* Nature Communications[J], 2017, 8: 15679.
- [26] SEKO A, MAEKAWA T, TSUDA K, *et al.* Physical Review B[J], 2014, 89(5): 054303.
- [27] YAN K Z, XU H B, SHEN G H, *et al.* Advances in Materials Science and Engineering[J], 2013, 2013: 597257.
- [28] WEN C, ZHANG Y, WANG C X, *et al.* Acta Materialia[J], 2019, 170: 109–117.
- [29] KOSTIUCHENKO T, KÖRMANN F, NEUGEBAUER J, *et al.* Npj Computational Materials[J], 2019, 5: 55.
- [30] SENKOV O N, WILKS G B, MIRACLE D B, *et al.* Intermetallics[J], 2010, 18(9): 1758–1765.
- [31] SENKOV O N, WILKS G B, SCOTT J M, *et al.* Intermetallics[J], 2011, 19(5): 698–706.
- [32] ZOU Y, MA H, SPOLENAK R. Nature Communications[J], 2015, 6: 7748.
- [33] WU Y D, CAI Y H, CHEN X H, *et al.* Materials & Design[J], 2015, 83: 651–660.
- [34] MELNICK A B, SOOLSHENKO V K. Journal of Alloys and Compounds[J], 2017, 694: 223–227.
- [35] HAN Z D, LUAN H W, LIU X, *et al.* Materials Science and Engineering: A[J], 2018, 712: 380–385.
- [36] POLETTI M G, MCCAUGHEY C M, FIORE G, *et al.* International Journal of Refractory Metals and Hard Materials[J], 2018, 76: 128–133.
- [37] KÖRMANN F, RUBAN A V, SLUITER M H F. Materials Research Letters[J], 2017, 5(1): 35–40.
- [38] 温晓灿, 张凡, 雷智锋, 等. 中国材料进展[J], 2019, 38(3): 242–249.
- WEN X C, ZHANG F, LEI Z F, *et al.* Materials China[J], 2019, 38(3): 242–249.
- [39] YANG T, ZHAO Y L, TONG Y, *et al.* Science[J], 2018, 362(6417): 933–937.
- [40] ZHAO Y L, YANG T, TONG Y, *et al.* Acta Materialia[J], 2017, 138: 72–82.
- [41] KUBE S A, SOHN S W, UHL D, *et al.* Acta Materialia[J], 2019, 166: 677–686.
- [42] BREIMAN L. Machine Learning[J], 2001, 45(1): 5–32.
- [43] DRUCKER H, BURGESS C J C, KAUFMAN L, *et al.* Support Vector Regression Machines[C]// Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1997: 155–161.
- [44] BRUCHER M, COURNAPEAU D. Scikit-Learn[R/OL]. (2017–07–18)[2019–02–28]. <https://scikit-learn.org/stable/>
- [45] GUO S, NG C P, LU J, *et al.* Journal of Applied Physics[J], 2011, 109(10): 103505.
- [46] YANG X, ZHANG Y. Materials Chemistry and Physics[J], 2012, 132(2/3): 233–238.
- [47] ZHANG Y, ZHOU Y J, LIN J P, *et al.* Advanced Engineering Materials[J], 2008, 10(6): 534–538.
- [48] BAIKOV A A. Database on Properties of Chemical Elements[R/OL]. (2018–03–20)[2019–02–28]. <http://phases.imet-db.ru/ELEMENTS/MAIN.ASPX>
- [49] TSAI M H, LI J H, FAN A C, *et al.* Scripta Materialia[J], 2017, 127: 6–9.
- [50] DOMINGUEZ L A, GOODALL R, TODD I. Materials Science and Technology[J], 2015, 31(10): 1201–1206.
- [51] LEONG Z Y, HUANG Y H, RUSSELL G, *et al.* Materials Chemistry and Physics[J], 2018, 210: 259–268.
- [52] 鲍美林, 乔珺威. 中国材料进展[J], 2018, 37(4): 24–32.
- BAO M L, QIAO J W. Materials China[J], 2018, 37(4): 24–32.
- [53] TIAN F Y, VARGA L K, CHEN N X, *et al.* Intermetallics[J], 2015, 58: 1–6.
- [54] DRAISSIA M, DEBILI M Y. Journal of Crystal Growth[J], 2004, 270(1/2): 250–254.

(本文为本刊约稿, 编辑 王 瑶)