

机器学习结合固溶强化模型预测高熵合金硬度*

张逸凡¹⁾ 任卫^{1)†} 王伟丽^{2)‡} 丁书剑²⁾ 李楠²⁾ 常亮¹⁾ 周倩¹⁾

1) (西安邮电大学理学院, 西安 710121)

2) (西北工业大学物理科学与技术学院, 西安 710072)

(2023 年 4 月 21 日收到; 2023 年 7 月 8 日收到修改稿)

第一性原理、热力学模拟等传统的材料计算方法在高熵合金的设计中多用于合金相的预测, 同时会耗费巨大的计算资源. 本文以性能为导向, 选用机器学习的算法建立了一个高熵合金硬度预测模型, 并将机器学习与固溶体强化的物理模型相结合, 使用遗传算法筛选出最具有代表性的 3 个特征参数, 利用这 3 个特征构建的随机森林模型, 其 R^2 达到了 0.9416, 对高熵合金的硬度取得了较好的预测效果. 本文选用的机器学习算法和 3 个材料特征在固溶体强化性质方面也有一定的预测效果. 针对随机森林可解释性较差的问题, 本文还利用 SHAP 可解释机器学习方法挖掘了机器学习模型的内在推理逻辑.

关键词: 高熵合金, 机器学习, 遗传算法, 固溶强化**PACS:** 07.05.Mh, 46.55.+d, 07.05.Kf, 61.66.Dk**DOI:** 10.7498/aps.72.20230646

1 引言

高熵合金 (HEA) 是由 4 种或 4 种以上金属元素按照等原子百分比或近似等原子百分比合成的合金材料^[1]. 高熵合金的各主元金属元素间发生复杂的相互作用而产生著名的四大效应^[2] (高熵效应、晶格畸变、迟滞扩散效应和“鸡尾酒”效应), 从而可能会使材料表现出极其优异的性能 (例如耐腐蚀性、高温热稳定性、力学性能、磁性等)^[3–5]. 然而, 由于高熵合金含有多种主元金属, 其成分的组合空间包含了 10^{78} 种组合方式^[6], 这就导致精确快速地筛选出具有优异目标性能的高熵合金成分是非常困难的.

第一性原理、热力学仿真等^[7–9] 传统的材料设计方法虽然能够加速新材料的发现, 但这些方法很难灵活地根据目标性能构建模型, 而且这些方法占

用了很大的计算资源, 计算成本很高, 然而预测精度一般. 随着人工智能的快速发展, 机器学习 (ML) 逐渐开始应用到材料科学中^[10–13]. 与传统材料计算方法相比, 机器学习通过大量的数据训练, 建立输入特征与目标性质之间的映射关系. 由于材料学科在漫长的发展进程中积累了大量的研究数据, 机器学习可以从这些数据中挖掘数据所蕴含的信息, 从而快速、精准地预测出材料的性质. Khakurel 等^[14] 选取了梯度提升算法来评估特征重要性, 有效地预测了难熔高熵合金的杨氏模量. Chang 等^[15] 使用成分加权和密度等材料特征建立了三层人工神经网络 (ANN) 来预测 AlCoCrFeMnNi 体系高熵合金的硬度. Bakr 等^[16] 利用 ANN 预测了高熵合金的硬度, 最终模型的决定系数达到了 0.88. 上述研究使用的模型虽然都具有一定的预测能力, 却难以兼顾模型的可解释性和预测精度的问题. Li 和 Guo^[17] 采用前向和后向特征选择得到的材料特征建立了

* 国家自然科学基金 (批准号: 51931005, 52171048, 51571163) 和陕西省创新产业链项目 (批准号: 2020ZDLGY12-02) 资助的课题.

† 通信作者. E-mail: renwei@xupt.edu.cn

‡ 通信作者. E-mail: wliwang@nwpu.edu.cn

一个支持向量机模型 (SVM) 用于预测高熵合金的合金相, 准确率超过了 90%. Xiong 等^[18] 利用前向特征选择法筛选的特征预测了合金相及相关力学性能. Lee 等^[19] 使用皮尔逊相关系数法 (PCC) 筛选特征并将神经网络算法集成获得了一个合金相分类器. Sun 等^[13] 利用 XGBoost 算法拟合相图计算辅助构建的 Ti-Zr-Nb-Ta 高熵合金硬度数据集, 通过特征重要性排名等特征选择方法, 揭示了预测该体系合金硬度的两个最重要的特征为 Ta 含量和熔点, 同时模型获得了 87.6% 的预测准确率. Wen 等^[20] 在模型构建的特征选择时利用皮尔逊相关系数法去除冗余特征, 随后通过将特征穷举完成特征选择, 并使得支持向量构建的硬度预测模型的精度有所提高. 最后联合支持向量机和效用函数成功搜索到了候选的高硬度 HEA. Li 等^[21] 将 Stacking 集成学习算法用于硬度预测, 成功降低了 HEA 硬度预测模型的预测误差. 然而, 上述研究使用的特征选择方法多为与模型无关的方法或贪心算法, 这些方法只覆盖了很少部分的特征组合, 甚至忽略了模型与特征间的关系, 这导致了建模时所使用特征组的质量相对较低. 此外, 在特征选择时上述研究大多针对单一特征集进行筛选, 在实际的特征选择过程中会忽略大量特征集之外的特征, 造成筛选出的特征仅有部分代表性. 而且由于高熵合金复杂的多主元结构及其形成机理, 构建一个完全包含所有影响硬度因素的特征集很困难. 而且上述研究筛选的特征都只针对某种单一 HEA 体系的性质进行预测是有效的, 而难以对其他 HEA 体系或 HEA 性质的性质进行预测. 此时, 构建一个科学合理的特征选择框架是至关重要的. 这将有利于构建一个适用于预测多种相关 HEA 性质的建模范特征集, 进而可以从机理上反映出 HEA 不同性能之间的关系. 因而, 在使用机器学习预测 HEA 性质的领域 (尤其是对 HEA 硬度的预测), 特征集的构建和特征选择的方法仍然具有改进空间. 此外, 由于机器学习多为黑盒模型, 所以预测模型的可解释性也至关重要, 这决定了是否可以进一步挖掘 HEA 形成的内部机制和预测模型的进一步优化.

本文首先建立了一个包含 19 个特征的高熵合金硬度数据集, 并利用该数据集选取建立模型的机器学习算法. 经过测试多种特征选择算法, 发现由遗传算法筛选出的候选特征质量较好. 同时, 为了克服候选特征迁移性差的问题, 结合传统固溶体强

化 (SSS) 物理模型, 对候选特征进行进一步筛选和优化. 最后, 采用了 SHAP 可解释机器学习方法^[22] 挖掘了输入特征对高熵合金硬度的影响机理.

2 建模条件

2.1 数据集与候选材料特征参数

本文选用文献^[20, 21] 使用的高熵合金材料硬度数据集, 该数据集包含了 Al, Co, Cr, Cu, Fe 和 Ni 六元高熵合金硬度样本, 计 205 条数据. 由于铸态条件下形成的相是稳定的, 所以合金的硬度数据均在铸态条件下测量^[23]. 数据集包含了一些数值异常的数据. 分析认为, 这些异常数据并不是因为测量误差导致的, 它们可能是由 HEA 固有的性质引起的数据差异, 应予以保留.

为了尽可能将与硬度相关的候选特征纳入特征集, 我们考察了曾广泛用于 HEA 性能预测的多个材料特征参数. 其中原子尺寸错配和模量错配对 HEA 的硬度有巨大的贡献^[20]. 基于此, 首先将原子半径误配 (δr)、原子堆叠失配因子 (γ)、杨氏模量 (E)、剪切模量 (G)、剪切模量误配 (δG)、晶格畸变能 (μ)、Peierls-Nabarro 因子 (F)、强化模型中的能量项 (A) 等与原子尺寸和模量相关的特征参数加入特征集中. 其次, 高熵合金的相与其硬度有着很强的关联性. Wang 等^[24] 发现功函数的六次方 (w^6) 与合金的屈服强度呈线性关系. Guo^[25] 回顾了根据经典 Hume Rothery 规则选取的 γ , Ω , Δ 等经验参数区分各类相的研究, 讨论了混合焓 (ΔH_{mix})、混合熵 (ΔS_{mix})、吉布斯自由能 (ΔG_{mix})、平均熔点 (T_m)、电负差 ($\Delta \chi$) 和价电子浓度 (VEC) 等参数对高熵合金相形成的影响. 这些参数也很重要, 需要被加入特征集. 此外, 巡回电子浓度 (e/a) 和内聚能 (E_c) 等与电子键合强度和电学性质有关的相稳定性参数也被加入特征集中. 由于 HEA 的硬度对组织变化很敏感, 还选取了一些与 HEA 力学性能相关的描述因子. 这样就得到了一个含有 19 个材料特征的 HEA 材料候选特征数据集. 相关参数及计算公式如表 1 所列, 其中 $r_{\min}$, $r_{\max}$ 代表 HEA 中最小和最大原子半径; c_i 代表各元素摩尔比; r 表示原子平均半径, $r = \sum_{i=1}^n c_i r_i$; R 为气体常数; H_{ij}^{mix} 表示第 i 和第 j 个元素之间的混合焓.

表 1 与高熵合金硬度相关的 19 个经验特征参数及其计算公式

Table 1. 19 empirical feature parameters related to the hardness of high entropy alloys and their calculation formulae.

材料特征	公式	材料特征	公式	材料特征	公式
$T_m, E_c, \text{VEC}, e/a, E, G$ (由 α 表示)	$\sum_{i=1}^n c_i \alpha_i$	ΔS_{mix}	$-R \sum_{i=1}^n c_i \ln(c_i)$	w^6	$\left(\sum_{i=1}^n c_i w_i\right)^6$
δG	$\sqrt{\sum_{i=1}^n c_i \left(1 - \frac{G_i}{G}\right)^2}$	ΔG_{mix}	$\Delta H_{\text{mix}} - T_m \Delta S_{\text{mix}}$	μ	$\frac{1}{2} E \delta r$
δr	$\sqrt{\sum_{i=1}^n c_i \left(1 - \frac{r_i}{r}\right)^2}$	$\Delta \chi$	$\sqrt{\sum_{i=1}^n c_i (\chi - \chi_i)^2}$	Ω	$T_m \frac{\Delta S_{\text{mix}}}{\Delta H_{\text{mix}}}$
γ	$\frac{1 - \sqrt{\frac{(r + r_{\min})^2 - r^2}{(r + r_{\min})^2}}}{1 - \sqrt{\frac{(r + r_{\max})^2 - r^2}{(r + r_{\max})^2}}}$	A	$G \delta r \frac{1 + \mu}{1 - \mu}$	A	$\frac{\Delta S_{\text{mix}}}{\delta r}$
ΔH_{mix}	$4 \sum_{i=1, j>i}^n c_i c_j H_{i-j}^{\text{mix}}$	F	$\frac{2G}{1 - \mu}$		

2.2 机器学习算法的选择

由于选择恰当的 ML 算法对于精准高效拟合 HEA 硬度数据具有重要的作用, 我们对适用于不同数据类型的 ML 算法及其特点进行了必要筛查.

集成学习算法是一种重要的 ML 算法. 早期, 因为其缺少可解释性而仅获有限的应用^[20,21,26]. 随着可解释 ML 的发展, 使用集成学习结合可解释 ML 的分析预测方法可以克服传统 ML 算法泛化性不足, 还可以兼顾模型的可解释性, 因而能挖掘大量有用信息. SVM 和 ANN 分别由于其核函数、支持向量机制和反向传播、激活函数机制使其具有强大的非线性映射能力, 因而广泛地应用在各种研究场景中. 此外, 基于线性回归的 ML 算法由于其强大的可解释性, 也被广泛用于建模中. Grinsztajn 等^[27]发现基于树的模型 (包括随机森林 (RF) 和 XGBoost 等算法) 在利用表格数据进行预测方面超过了深度残差网络等其他深度学习模型. 由于本文使用的 HEA 硬度数据集样本规模小, 数据不均匀, 基于树的模型可能更适合该数据集. 虽然上述模型都可以确立特征与目标值之间的隐式关系, 不同的 ML 模型对数据采用不同的处理手段, 所获得的模型可解释性是不同的, 在建模中要根据不同的目标值选取合适的 ML 模型. 本文在建模中使用了 Python 的 sklearn, pandas, numpy 等常用数据分析库.

最后, 在利用 HEA 数据集构建 ML 模型之前, 还要考虑如何降低模型过拟合或欠拟合所带来的风险. 传统的按固定比例划分数据集来评价模型精

度的方法会因为数据不平衡而导致模型泛化性差. 所以本文在建立 ML 模型过程中始终将交叉验证的方法应用于 ML 模型选择、ML 参数调优、ML 预测结果评估等各个阶段, 从而保证了建模结果的科学性和可信性. 各阶段使用的评价依据均为方误差

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

和决定系数

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

3 建模和结果

3.1 基准机器学习算法

为选出最合适的基准算法, 本文使用了具有优异外推能力的线性算法 (Ridge 和 Lasso), SVM, 具有单隐藏层的 ANN, 性能优异但外推能力较差的集成学习算法 (如基于 Bagging 的代表算法 RF 和基于 Boosting^[28]的代表算法 XGBoost). 首先, 将前述 19 个候选材料特征作为输入参数, 将 HEA 硬度值作为输出值, 依次对上述 ML 算法进行训练. 为确保发挥出各个 ML 算法的特点和优势, 在对 ML 算法进行训练时, 将网格搜索法和十折交叉验证法 (10-fold) 相结合搜索使模型 RMSE 最小的超参数. 经过寻优搜索和交叉验证, 在 SVM 的 3 种核函数中选择了非线性映射能力最好的高斯核函

表 2 不同机器学习模型搜索的超参数结果
Table 2. Hyperparametric search results for different machine learning models.

算法	超参数
SVM-rbf	$\gamma = 1 \times 10^{-7}$, $C = 200$
RF	$\max_depth = 6$, $\min_samples_leaf = 1$, $\min_samples_split = 2$, $n_estimators = 50$
XGBoost	$\gamma = 0.1$, $learning_rate = 0.1$, $\max_depth = 12$, $n_estimators = 100$, $reg_alpha = 0$, $reg_lambda = 0.5$
ANN	$\max_iter = 210000$, $hidden_layer_sizes = 16$, $solver = 'adam'$, $activation = 'relu'$, $\alpha = 0.01$
Lasso	$\alpha = 1$, $\max_iter = 10000$
Ridge	$\alpha = 0.1$, $\max_iter = 10000$

数 (SVM-rbf). ANN 在使用 Adam 作为反向传播优化器和线性修正函数作为激活函数时取得了最优的结果. 机器学习模型搜索的超参数结果如表 2 所列, 所有超参数的相关解释详见 sklearn. 此外, 为了增强模型的物理可解释性, 并没有对特征集进行数据标准化或数据降维等数据预处理操作 [29,30].

将搜索的超参数应用到 ML 算法上, 通过 10-fold 验证的方法评估了前述 ML 算法的 RMSE 和 R^2 , 结果如图 1 所示.

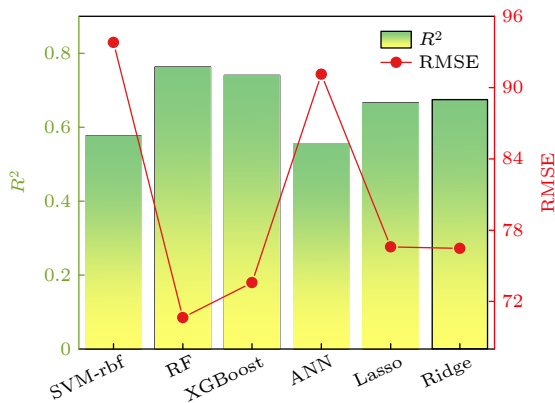


图 1 6 种机器学习算法对数据集的拟合结果
Fig. 1. Fitting results of six machine learning algorithms to the dataset.

从图 1 可以看出, RF 不仅取得了最高的 R^2 , 并且有着最小的 RMSE, 这表明了 RF 在该数据集上具有最佳拟合效果. 所以将 RF 作为接下来进行特征选择和模型建立的基准算法.

3.2 优化特征组筛选及特征解释

从候选特征集筛选出优化的特征组合, 经过模型训练, 能够使数据拟合的误差达到最小. 由于不同数据集的样本分布不同, 所以在对该数据集先验知识不足的情况下, 需要利用多种与 ML 算法相关的特征选择方法主动选择适合该 ML 算法的特征

组合, 从而更好地解释目标属性. 本文分别使用了包裹法和嵌入法进行特征筛选. 包裹法通过使用特征搜索策略修改特征组合, 以此来选择出优化特征子集. 该方法主要包括遗传算法 (GA)、序列前向选择 (SFS)、序列后向选择 (SBS)、递归特征消除 (RFE) 等方法. 嵌入法利用前述 RF 算法学习器对特征重要性进行评估, 依次从特征集中剔除不重要特征, 以此筛选出表现最好的特征组. 其中, GA 是一种通过模拟自然选择、遗传和变异等生物进化过程来寻找最优解的最优化算法. GA 通过初始化种群、评估适应度、选择、交叉和变异等步骤来寻找最优问题的最优解. 具体地, 将 GA 运用到特征选择时, 本文将随机森林模型在十折交叉验证法下的 RMSE 作为适应度用于评估每个特征集合的优良程度; 初始化种群即为所有可能的特征集合的集合; 在执行选择操作时将优秀的特征集合复制到下一代来保留优秀的基因, 同时引入新的变异来增加种群的多样性; 在执行交叉操作时, 将两个特征集合的某些部分进行交换, 以产生新的特征集合. 交叉操作可以促进基因的流动和交换, 从而增加种群的多样性; 在执行变异操作时, 通过添加或删除某些特征来增加特征集合的随机性, 以增加种群的多样性. 在执行 GA 时使用了 python 中的 genetic_selection 库, 将最小化随机森林模型在十折交叉验证法下的 RMSE 为目标, 利用 GA 对特征集进行全局特征搜索, 最后通过执行上述优化步骤筛选出最优的特征组. 对于 RFE, SFS, SBS 以及基于 RF 的包裹法, 利用 python 中的 sklearn 库, 测试了在各个特征选择方法下保留不同特征数量时的特征选择结果. 如图 2 所示, 对除全局优化算法的 GA 之外的其他 4 种特征选择方法进行测试. 在使用不同特征选择算法时, 设定保留的特征数量分别为 1 到 19, 选择出不同特征选择算法在保留不同特征数量下的特征集. 随后, 通过对比各个特征选择算

法在保留不同特征数量下的 RMSE, 选出各个特征选择算法下的最优特征集. 如表 3 所列, RFE 和 RF 分别筛选出了含有 13 个特征的优化特征组, 虽然它们使用了较多的材料特征而对 HEA 硬度预测的表现则较差. SBS 和 SFS 算法筛选出了含有 7 个特征的优化特征组, 其 RMSE 约为 67, 预测表现略高于 RFE 和 RF. 而 GA 筛选出了含有 8 个特征的优化特征组, 其 RMSE 仅为 64.09. 预测表现明显优于其他特征选择算法. 这可能是因为 GA 是一个全局搜索算法, 其遍历的特征组合更加全面. 因而本文使用 GA 进行特征选择.

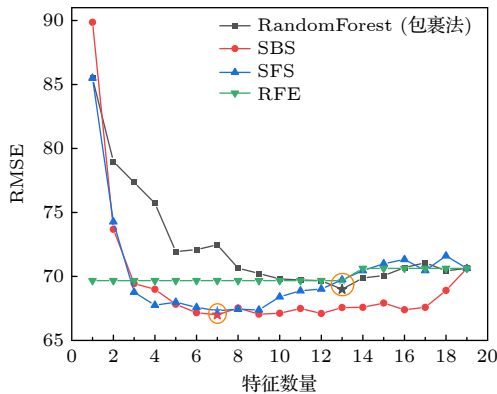


图 2 SBS, SFS, RF, RFE 算法在不同特征数下选择的最佳特征的 RMSE, 曲线中的星号代表了当前特征选择方法选择的最优特征组所包含的特征数

Fig. 2. Different number of features selected by SBS, SFS, RF, RFE algorithm vs. their RMSE performances under 10 fold. The asterisks in the curves represent the number of features contained in the optimal feature group selected by the current feature selection method.

表 3 不同特征选择方法筛选的优化特征组及 RMSE 值
Table 3. Optimized feature sets screened by different feature selection algorithms and their RMSE values.

算法	优化特征组	RMSE
GA	$\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G$	64.09
SFS	$\delta r, E_c, \text{VEC}, \Delta H_{\text{mix}}, \Omega, E, G$	67.32
SBS	$\Delta\chi, E_c, \text{VEC}, A, w, F, \delta G$	67.00
RFE	$\delta r, E_c, \text{VEC}, \Delta S_{\text{mix}}, \Omega, A, E, \mu, w, G, F, A, \delta G$	69.67
RF	$\delta r, \text{VEC}, F, A, w, \delta G, \mu, G, A, E_c, \Omega, \Delta S_{\text{mix}}, \Delta H_{\text{mix}}$	68.99

为了克服集成学习可解释性差的问题, 采用 Lundberg 和 Lee [31] 提出的沙普利加和解释 (SHAP) 方法. SHAP 是一种解释 ML 模型输出的博弈论方法, 通过计算各个特征对预测结果的边际贡献, 完

成对黑盒模型局部或全局的分析. 如图 3 所示, 利用 SHAP 方法解释 GA 选择的优化特征组特征对 HEA 硬度的影响. 根据文献 [25], VEC, $\Delta\chi$ 等材料特征参数能影响 HEA 形成 FCC 相和 BCC 相的稳定性: 当 $\text{VEC} < 6.87$ 时 HEA 倾向于生成 BCC 相, 当 $\text{VEC} > 8$ 时倾向于生成 FCC 相, 当 $6.87 < \text{VEC} < 8$ 时则倾向于生成混合相. 由于 BCC 相的晶格畸变比 FCC 相更严重, BCC 相的硬度一般要大于 FCC 相. 因而, 如果 HEA 材料的 VEC 值较小, 表明其更倾向于生成 BCC 相, 也具有更大的硬度, 这与 SHAP 值的分析结果是一致的. F 描述了晶体位错滑移所需要的点阵阻力. 对于 FCC 相来说, 由于其具有较大位错宽度, 晶体位错滑移所产生的阻力也就较小, 屈服应力也较低, F 也相对较小. 相比之下, BCC 相具有较小的位错宽度, 晶体位错滑移产生的阻力比较大, F 和屈服应力都比较大 [32]. HEA 的屈服强度和维氏硬度有大致的线性对应关系 [33], 表明 F 与 HEA 硬度也有着正相关的关系. 传统固溶强化理论认为模量错配对 SSS 具有重要的促进作用, δG 反映了元素原子键合的强度. 而 HEA 的力学性能与原子键合强度密切相关 [34], 因而 δG 对 HEA 硬度具有积极促进作用. Wang 等 [35] 提出了一个反映原子尺寸差异影响 HEA 固溶度程度的参数 γ , γ 可以很好地划分不同 HEA 相的形成范围. 当 $\gamma < 1.175$ 时 HEA 倾向于生成固溶相, 当 $\gamma > 1.175$ 时, HEA 更倾向于生成金属间化合物 (IM). IM 相对于固溶相 (SSP) 来说, 其硬度更大, 脆性也更大. 同样地, 根据 Hume-Rothery 准则, $\Delta\chi$ 较大时, HEA 体系中具有高电正性的元素原子容易失去核外电子, 高电负性元素原子则容易得电子, 故高电负性元素吸引共用电子的能力也会变大, 此时 HEA 更易形成 IM, 从而提高 HEA 硬度. Yang 和 Zhang [36] 提出的热力学参数 Ω 反映了熵和焓之间的竞争. 考虑 ΔH_{mix} 和 ΔS_{mix} 等热力学参数对 HEA 相形成的影响, 其中 ΔH_{mix} 阻碍固溶体的形成, ΔS_{mix} 促使形成固溶体, 当 Ω 参数较小时, 说明 ΔH_{mix} 在 HEA 体系固溶体形成中处于主导地位, 此时不易形成固溶体, 硬度较大的 IM 将优先形成; 反之, 当 Ω 参数较大时, ΔS_{mix} 在 HEA 体系固溶体形成中处于主导地位, 容易生成固溶体. 经过 SHAP 方法的分析, GA 选出的优化特征组所包含的特征参数 [$\gamma, \Delta\chi, \delta G$] 在 SHAP 排序中靠前, 表明这些特征在 ML 建模中很重要. 同时

$[\delta G, E]$ 参数也与模量相关, 而模量参数在传统物理计算模型中占据了非常重要的部分. 这些结果表明通过 GA 筛选出的材料特征具有重要物理意义, 对于进一步改进力学性能预测模型具有很好的指导作用.

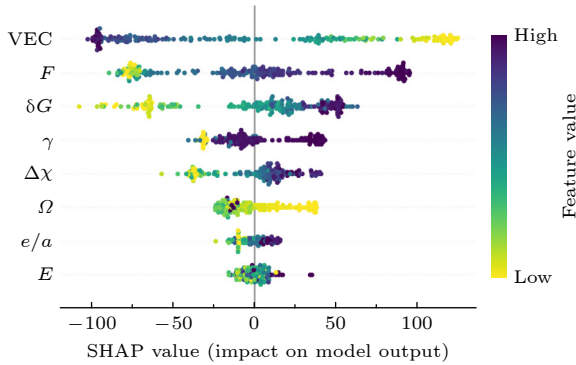


图 3 遗传算法所选优化特征组 8 种特征的 SHAP 分析, 8 种特征由上到下重要性依次降低, 各个散点根据 SHAP 值的正负反映了该特征的大小对当前样本点硬度的促进或削弱作用

Fig. 3. SHAP analysis of the eight features of the optimized feature set selected by the genetic algorithm. The eight features decrease in importance from top to bottom. Each scatter reflects the promoting or weakening effect of the size of the feature on the hardness of the current sample point according to the positive or negative SHAP value.

3.3 基于固溶强化理论优化输入特征

为分析 GA 所选择的 8 个特征间是否存在冗余特征, 明确是否需要进一步优化特征集, 计算了各个特征以及 HEA 硬度之间的 PCC. 图 4(a) 的子图是使用基准算法 RF 评估的各个特征的重要性排序, 其中 VEC, F , δG , $\Delta\chi$ 等参数与 PCC 得到的特征和硬度间的相关性基本一致. 将 $|PCC| > 0.8$ 的特征视为高相关特征. 如图 4(a) 所示, $[\gamma, F]$, $[\Delta\chi, F]$, $[\text{VEC}, e/a, \delta G]$ 这些特征组合相关性相对较高. 对于这些特征组合, 不能简单通过评估特征重要性来删除不重要特征, 因为这样做可能忽略了各个特征对 HEA 硬度的协同促进作用. 此外, PCC 是基于变量之间的线性相关性来衡量它们之间的关联程度. 然而, 如果特征之间存在非线性关系, PCC 可能无法准确反映它们之间的相关性. HEA 复杂的形成机制蕴含着复杂的非线性关系. 利用 PCC 筛选特征, 可能会忽略这些关系, 从而漏选重要特征. 其次, PCC 仅考虑特征之间的两两关系, 并不能全面捕捉多个特征之间的复杂关系.

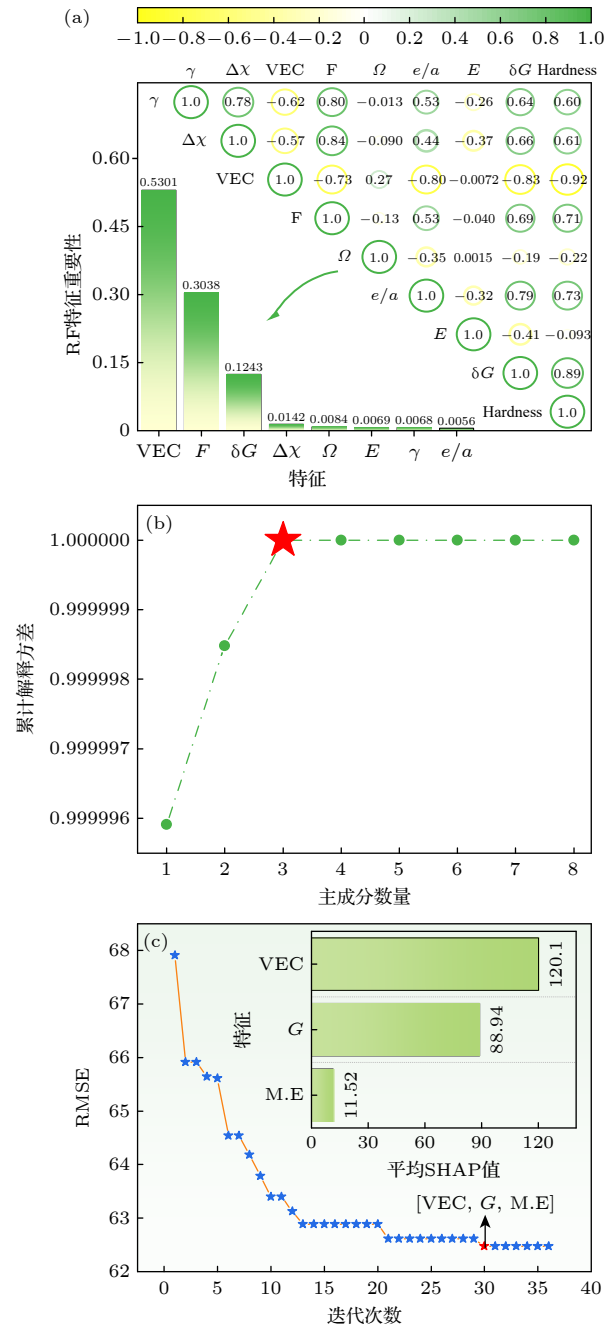


图 4 (a) 遗传算法所选特征的 PCC 热图, 子图为遗传算法所选特征的 RF 重要性评估排序; (b) 主成分分析法计算优化特征组 $[\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G]$ 不同主成分分数的累计方差贡献率; (c) 新构建的特征集进行 GA 特征选择的迭代过程, 子图为 GA 选择特征的 SHAP 重要性排序

Fig. 4. (a) PCC heat map of the features selected by the genetic algorithm, with subplots for the RF importance assessment ranking of the features selected by the genetic algorithm; (b) the cumulative variance contribution of different principal component scores of the optimized feature set $[\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G]$ calculated by principal component analysis; (c) iterative process of GA feature selection for the newly constructed feature set, and the subplot is the SHAP importance ranking of the GA selected features.

在复杂的 HEA 材料数据中, 多个特征之间可能存在更高阶的相互作用或非线性关系, 这些关系无法通过皮尔逊相关系数来准确表示. 最后, PCC 只考虑了特征之间的相关性, 而忽略了特征与目标变量之间的关联性, 这可能会删除一些对于目标值预测比较重要的特征. 图 4(b) 对该特征组进行主成分分析, 发现在 8 个特征中, 只需提取 3 个主成分即可保留特征集所有的信息. 这为我们选取更加优化的输入特征提供了一个思路. 同时, 这也意味着特征集仍具有改进的空间. 为了提升模型精度, 令模型更具可解释性, 需要依据当前特征选择的结果, 进一步扩充特征集, 选择更加具有代表性的特征, 进而优化建模范征组.

HEAs 的强化机制主要由 SSS 导致, 其中屈服强度和维氏硬度正相关, 且屈服强度约为维氏硬度的 9.81/3 倍, 所以 SSS 对 HEA 维氏硬度的提升很重要.

一般认为, HEAs 的 SSS 主要由溶质原子引起的晶格畸变和滑动位错引起, 其中金属元素的尺寸错配和模量错配引起的晶格畸变在合金强化中占据重要地位. 经典的 Labusch 模型^[37]体现了尺寸误配和模量误配对 SSS 的影响. 传统的 SSS 数学模型, 大多是基于 Labusch 模型进行改进. Thirathipiwat 等^[38]发现高晶格畸变引起显微硬度的变化与 δr 正相关; Ma 和 Wu^[39]也发现 δr 有利于导致位错线形成波浪形构型, 这为 HEA 提供了显著的 SSS 效果. Toda-Caraballo 和 Rivera-Díaz-del-Castillo^[40]利用 Gypen 模型将 Labusch 模型从二元合金推广到稀释的多组分合金, 通过计算原子间距变化, 量化各组分引起的晶格畸变对 SSS 的贡献来估计合金的 SSS 程度. 所以原子尺寸失配与模量失配对 SSS 有着重要影响. Toda Caraballo 提出的 SSS 模型等式可表示为^[34]

$$\Delta\sigma_{\text{SSS}} = Z\xi G\delta, \quad (1)$$

其中 ξ 为 SSS 模型的结构因子, BCC 相为 4, FCC 相为 1; δ 为模型的 SSS 强化因子, 与原子尺寸误配相关; $\Delta\sigma_{\text{SSS}}$ 为量化固溶强化程度的参数; Z 为固溶强化因子.

基于上述理论, 我们发现在 GA 所选的 8 个特征中, HEA 的硬度与模量, 原子半径和电负差的误配有关. 这也暗示了此类特征与 HEA 硬度增强有关. 以此为基础, 将 3.2 节中筛选出的 E , G 以及体

积模量 (K) 作为原始数据, 通过 (2)—(6) 式计算 5 个与误配相关的特征:

$$\delta d = \sqrt{\sum_{i=1}^n c_i \left(1 - \frac{d_i}{d}\right)^2}, \quad (2)$$

$$\Delta d = \sqrt{\sum_{i=1}^n c_i (d - d_i)^2}, \quad (3)$$

$$D.d = \sum_{i=1}^n \sum_{j=1, i \neq j}^n c_i c_j |d_i - d_j|, \quad (4)$$

$$\eta.d = \sum_{i=1}^n \frac{c_i \frac{2(d_i - d)}{d_i + d}}{1 + 0.5 \left| c_i \frac{2(d_i - d)}{d_i + d} \right|}, \quad (5)$$

$$M.d = \text{Max} \left[c_i \left(1 - \frac{d_i}{d}\right)^2 \right] - \text{Min} \left[c_i \left(1 - \frac{d_i}{d}\right)^2 \right], \quad (6)$$

其中 d 分别代表杨氏模量、剪切模量、体积模量、原子半径和电负性等参数, c_i 代表 HEA 各元素的摩尔比, d_i 代表 HEA 各元素的参数 α_i 值. 由于 γ 参数与金属原子尺寸误配相关, 所以将原子半径同样利用 (2)—(6) 式的方法扩充尺寸误配特征. 将重新计算的特征与 3.2 节中筛选的特征组 $[\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G]$ 整合到一起, 组成了含有 35 个特征参数的扩充特征集, 然后通过 GA 重新进行特征选择.

如图 4(c) 所示, GA 在迭代到第 30 代时, 收敛到最优解, 此时搜寻到的简版优化特征组为 $[\text{VEC}, G, \text{M.E}]$. 相对于 GA 最初筛选的包含 8 个特征的优化特征组, 简版优化特征组仅用 3 个参数来构建 ML 模型, 并且由于使用 G 和 M.E 这两个特征取代其他大量的特征, 极大减少了冗余特征, 降低了模型复杂度.

在对 ML 模型进行评估的时候, 不仅要评估模型泛化性, 还要评估模型的外推能力. 10-fold 是将数据集平均分为 10 份, 评估 ML 算法在其中 9 份数据集上的训练结果是否可以迁移到剩余的 1 份数据集上. 该方法可以用来评估 ML 模型的泛化能力. 由于采用的材料硬度数据具有稀疏性, 可能由于数据集划分不当, 导致模型训练信息遗漏, 使模型在不同测试集上的表现差距较大, 导致模型缺乏外推能力. 因而可以采用留一交叉验证 (LOOCV) 评价模型的外推能力. 具体作法是分别将简版优化

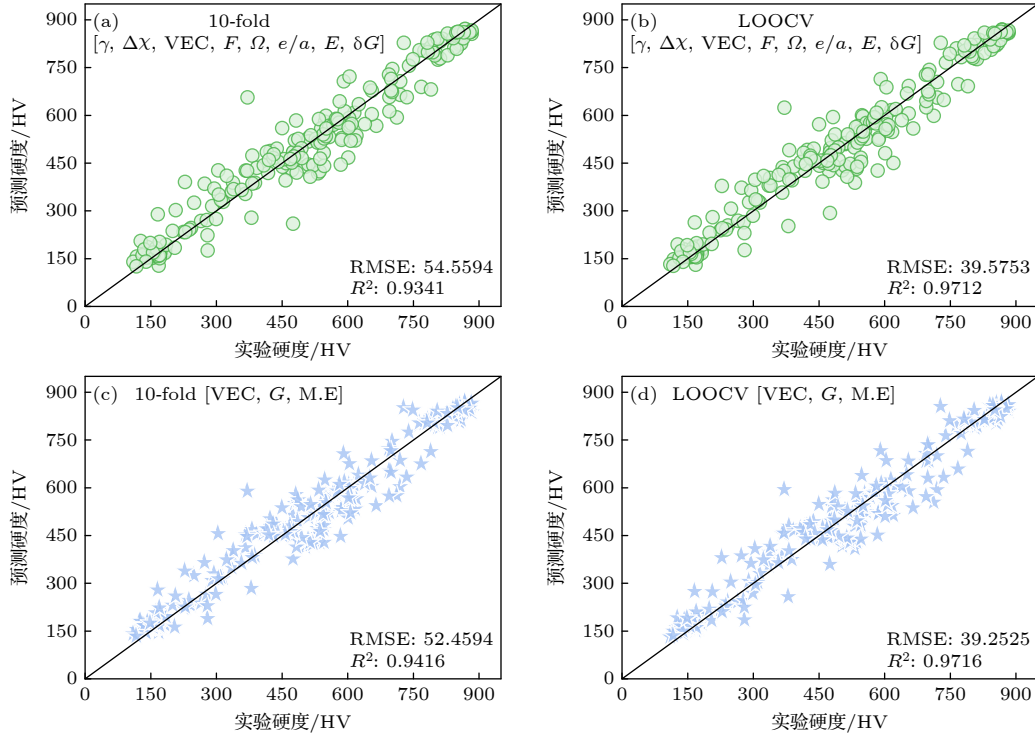


图 5 (a), (c) 在十折交叉验证下的模型拟合结果以及 (b), (d) 在 LOOCV 下的模型拟合结果, 其中 (a), (b) 优化特征组 $[\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G]$ 作为 RF 输入特征; (c), (d) 简版优化特征组 $[\text{VEC}, G, \text{M.E}]$ 作为 RF 输入特征

Fig. 5. (a), (c) Model fit results under 10-fold cross-validation and (b), (d) model fit results under LOOCV: (a), (b) Optimized feature set $[\gamma, \Delta\chi, \text{VEC}, F, \Omega, e/a, E, \delta G]$ as RF input features; (c), (d) the short version of the optimized feature set $[\text{VEC}, G, \text{M.E}]$ as RF input features.

特征组和优化特征组作为 RF 的输入, 利用贝叶斯优化方法优化 RF 超参数提高模型预测性能. 如图 5 所示, 虽然简版优化特征组去除了多数输入特征参数, 但是模型在 10-fold 和 LOOCV 情况下的 RMSE 和 R^2 相较于优化特征组仍然有一定程度的进步. 为了进一步分析 2.1 节所述异常值对本文建模的影响, 使用了孤立森林算法对数据集进行异常值检测. 孤立森林算法是一种无监督的离群点检测算法. 该算法的优点包括能够快速准确地识别异常值点, 并且内存使用率低. 此外, 由于该算法建立在基于树的结构上, 因此实现简单, 是异常检测的有效工具. 如图 6(c) 所示, 使用孤立森林对数据集的样本进行评分, 得分小于零则视为离群点, 最后发现了 11 个离群点. 将这 11 个离群点从数据集剔除后, 分别使用 LOOCV 和 10-fold 评估模型的 R^2 和 RMSE. 如图 6(a) 和图 6(b) 以及图 5(c) 和图 5(d) 所示, 去除异常值后的 R^2 和 RMSE 相较于去除前, 精度有所下降. 这可能是由于我们在所使用的数据集均来源于真实世界, 并且通过科学合理的方法测量, 所以异常值点是由测量误差导致的概率比较低. 异常值点在真实世界是客观存在的, 所以一

些异常点的情况必须被考虑到其中, 这可以建立模型和真实世界的联系. 同时, 如果在预测新样本时, 该样本恰好与以前剔除的异常值点类型相同, 此时预测偏差可能会急剧上升. 所以在本文中将会保留异常值. 另一方面, 由于 LOOCV 对异常值点的存在非常敏感, 所以 LOOCV 的评估结果对于模型离群点的实际影响可以作为参考. 如图 6(b) 和图 5(d) 所示, 在 LOOCV 的测试下, 并没有出现很明显的离群点. 这也证明了少量的异常值点对建模的影响并不大.

为了进一步探索特征集 $[\text{VEC}, G, \text{M.E}]$ 在 SSS 的应用, 对 SSS 物理模型进行改进. 首先经验参数 VEC 与 ξ 的作用类似, 都能区分具有 FCC 和 BCC 结构固溶体相的 HEA. 但 ξ 参数划分不同相的边界更加清晰. Wen 等^[34] 使用 ξ 代替 VEC 训练 ML 模型, 获得了更好的 SSS 预测结果. 针对 Toda-Caraballo 和 Rivera-Díaz-del-Castillo^[40] 提出的 SSS 物理模型, 将特征集 $[\text{VEC}, G, \text{M.E}]$ 中的 VEC 替换为 ξ , 得到了该 SSS 模型的另外一种表现形式:

$$\Delta\sigma_{\text{SSS}} = Z\xi\text{GM.E.} \quad (7)$$

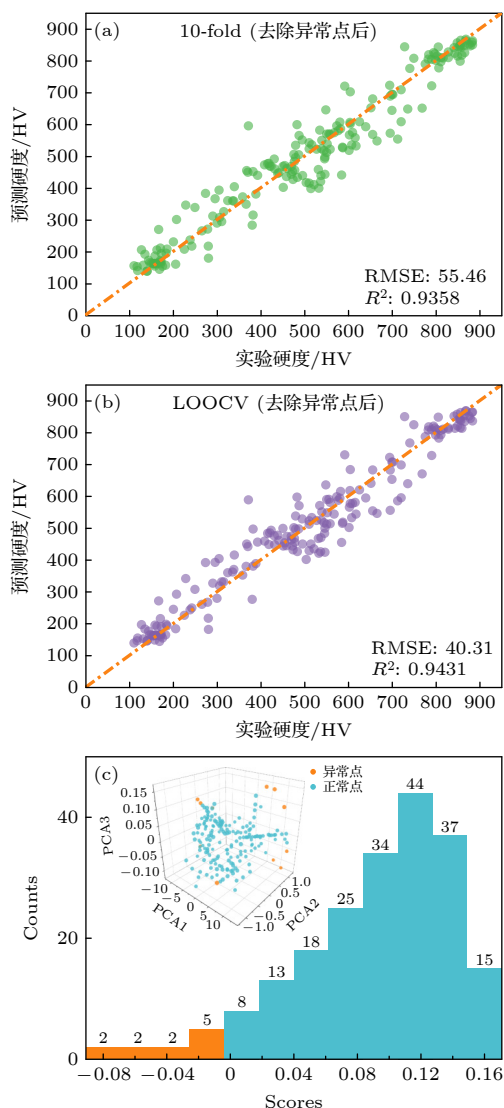


图6 数据集去除异常值后的拟合图 (a) 使用了 10-fold 评估; (b) 使用了 LOOCV 评估; (c) 主图为异常值得分结果, Scores < 0 视为离群点; 利用孤立森林对 205 个高熵合金样本进行异常值检测, 子图为利用主成分分析法降维后的异常值检测可视化结果

Fig. 6. Fitted plots of the dataset after removing outliers: (a) 10-fold is used; (b) LOOCV is used; (c) the outlier score histogram (the orange points being outlier points when scores < 0). The outlier detection is carried out for 205 high-entropy alloy samples by using isolated forest. The inset 3D figure shows the visualization results of the outlier detection after the dimensionality reduction by using principal component analysis.

为了验证 (7) 式的准确性, 采用文献 [34] 收集的 162 条铸态 HEA 的 SSS 贡献 ($\Delta\sigma_{SSC}$) 数据集, 该数据集中所包含的金属元素不仅有 3d 过渡金属元素 (Co, Cr, Cu, Fe, Ni, Mn, Ti, V), 还包含 Zr, Hf, Mo, Nb, Ta, W, Al 等难熔金属元素. 为了更好地估计 SSS 对 HEA 硬度的贡献, 该数据集中只保

留了 BCC 和 FCC 的单相固溶体数据, 这样处理可以削弱除 SSS 外的其他强化效果的干扰. 如图 7 所示, 将 $[\xi, G, M.E]$ 作为 RF 算法输入, $\Delta\sigma_{SSC}$ 作为输出, 经过贝叶斯优化方法对 RF 进行超参数优化, 最终在 10-fold 的情况下 RMSE 和 R^2 分别为 542.3691 和 0.8811.

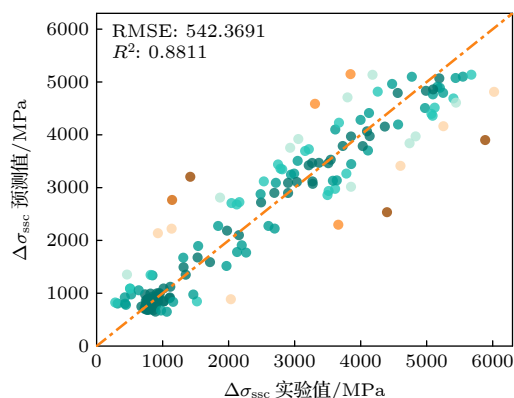


图7 以 $[\xi, G, M.E]$ 作为 RF 输入特征, $\Delta\sigma_{SSC}$ 作为目标值, 在十折交叉验证下的评估结果

Fig. 7. Evaluation results with $[\xi, G, M.E]$ as the RF input features and $\Delta\sigma_{SSC}$ as the target values under 10-fold cross-validation.

4 结 论

本文首先使用集成学习算法对一个包含 19 个候选特征的高熵合金硬度数据集进行训练、测试和评估: 从多种特征选择算法中筛选出遗传算法对 19 个候选特征进行筛选, 获得包含 8 个特征的优化特征组; 然后结合两阶段的特征选择方法, 利用传统固溶强化模型优化建模特征, 最终筛选出包含 3 个材料特征的简版最优特征组合 $[VEC, G, M.E]$; 利用这 3 个特征建立的 RF 模型使得模型在十折交叉验证法下的 R^2 达到了 0.9416, RMSE 达到了 52.4594. 基于该特征组合建立的模型对于固溶强化的预测也具有一定适应性, 在预测固溶体强化的贡献时 R^2 达到了 0.8811, 这表明该模型对高熵合金力学性质的预测可能也有好的迁移效果. 最后, 本文使用可解释机器学习挖掘 HEA 硬度数据的隐含信息, 初步揭示了一些重要材料特征对 HEA 硬度的影响机理.

参考文献

- [1] Wu P F, Gan K F, Yan D S, Fu Z H, Li Z M 2021 *Corros. Sci.* **183** 109341

- [2] Ranganathan S 2003 *Curr. Sci.* **85** 1404
- [3] Wu Y D, Cai Y H, Wang T, Si J J, Zhu J, Wang Y D, Hui X D 2014 *Mater. Lett.* **130** 277
- [4] Yu Y, Wang J, Li J S, Kou H C, Duan H T, Li J, Liu W M 2015 *Tribol. Int.* **92** 203
- [5] Cheng P, Zhao Y H, Xu X T, Wang S, Sun Y Y, Hou H 2020 *Mater. Sci. Eng. A* **772** 138681
- [6] Beniwal D, Singh P, Gupta S, Kramer M J, Johnson D D, Ray P K 2022 *npj Comput. Mater.* **8** 153
- [7] Jiang C, Uberuaga B P 2016 *Phys. Rev. Lett.* **116** 105501
- [8] Li Z Q, Tan X Y, Duan X L, Zhang J Y, Yang J Y 2022 *Acta Phys. Sin.* **71** 247803 (in Chinese) [李志强, 谭晓瑜, 段忻磊, 张敬义, 杨家跃 2022 物理学报 **71** 247803]
- [9] Ren X L, Zhang W W, Wu X Y, Wu L, Wang Y X 2020 *Acta Phys. Sin.* **69** 046102 (in Chinese) [任县利, 张伟伟, 伍晓勇, 吴璐, 王月霞 2020 物理学报 **69** 046102]
- [10] Yang Z X, Gao Z R, Sun X F, Cai H L, Zhang F M, Wu X S 2019 *Acta Phys. Sin.* **68** 210502 (in Chinese) [杨自欣, 高章然, 孙晓帆, 蔡宏灵, 张凤鸣, 吴小山 2019 物理学报 **68** 210502]
- [11] Kou W B, Dong H, Zou M Q, Han J Y, Jia X X 2021 *Acta Phys. Sin.* **70** 030701 (in Chinese) [寇雯博, 董灏, 邹岷强, 韩均言, 贾西西 2021 物理学报 **70** 030701]
- [12] Li W, Long L C, Liu J Y, Yang Y 2022 *Acta Phys. Sin.* **71** 060202 (in Chinese) [黎威, 龙连春, 刘静毅, 杨洋 2022 物理学报 **71** 060202]
- [13] Sun Y, Lu Z C, Liu X J, Du Q, Xie H M, Lv J C, Song R X, Wu Y, Wang H, Jiang S H, Lu Z P 2021 *Appl. Phys. Lett.* **119** 201905
- [14] Khakurel H, Taufique M F N, Roy A, Balasubramanian G, Ouyang G, Cui J, Johnson D D, Devanathan R 2021 *Sci. Rep.* **11** 17149
- [15] Chang Y J, Jui C Y, Lee W J, Yeh A C 2019 *JOM* **71** 3433
- [16] Bakr M, Syarif J, Hashem I A T 2022 *Mater. Today Commun.* **31** 103407
- [17] Li Y, Guo W L 2019 *Phys. Rev. Mater.* **3** 095005
- [18] Xiong J, Shi S Q, Zhang T Y 2021 *J. Mater. Sci. Technol.* **87** 133
- [19] Lee K, Ayyasamy M V, Delsa P, Hartnett T Q, Balachandran P V 2022 *npj Comput. Mater.* **8** 25
- [20] Wen C, Zhang Y, Wang C X, Xue D Z, Bai Y, Antonov S, Dai L H, Lookman T, Su Y J 2019 *Acta Mater.* **170** 109
- [21] Li S, Li S, Liu D R, Zou R, Yang Z Y 2022 *Comput. Mater. Sci.* **205** 111185
- [22] KUSDHANY M, LYTH S M 2021 *Carbon* **179** 190
- [23] Chang H, Tao Y, Liaw P K, Ren J 2022 *J. Alloys Compd.* **921** 166149
- [24] Wang W Y, Shang S L, Wang Y, Han F, Darling K A, Wu Y, Xie X, Senkov O N, Li J, Hui X D, Dahmen K A, Liaw P K, Kecskes L J, Liu Z K 2017 *npj Comput. Mater.* **3** 23
- [25] Guo S 2015 *Mater. Sci. Technol.* **31** 1223
- [26] Yang C, Ren C, Jia Y, Wang G, Li M, Lu W 2022 *Acta Mater.* **222** 117431
- [27] Grinsztajn L, Oyallon E, Varoquaux G 2022 *NeurIPS 2022 Datasets and Benchmarks Track* New Orleans, United States, November 28, 2022 p507
- [28] Zhang Y F, Ren W, Wang W L, Li N, Zhang Y X, Li X M, Li W H 2023 *J. Alloys Compd.* **945** 169329
- [29] Zhang Y, Wen C, Wang C X, Antonov S, Xue D Z, Bai Y, Su Y J 2020 *Acta Mater.* **185** 528
- [30] Zhang L, Chen H M, Tao X M, Cai H G, Liu J N, Ouyang Y F, Peng Q, Du Y 2020 *Mater. Des.* **193** 108835
- [31] Lundberg S, Lee S I 2017 *Proceedings of the 31st International Conference on Neural Information Processing Systems*, United States, December 3, 2017 p4768
- [32] Hu G X, Cai X, Rong Y H 2010 *Material Science Foundation* (Shanghai: Shanghai Jiao Tong University Press) p177 (in Chinese) [胡赓祥, 蔡珣, 戎咏华 2010 材料科学基础 (上海: 上海交通大学出版社) 第 177 页]
- [33] Huang X Y, Jin C, Zhang C, Zhang H, Fu H W 2021 *Mater. Des.* **211** 110177
- [34] Wen C, Wang C X, Zhang Y, Antonov S, Xue D Z, Lookman T, Su Y J 2021 *Acta Mater.* **212** 116917
- [35] Wang Z, Huang Y, Yang Y, Wang J, Liu C T 2015 *Scr. Mater.* **94** 28
- [36] Yang X, Zhang Y 2012 *Mater. Chem. Phys.* **132** 233
- [37] Labusch R 1970 *Phys. Status Solidi B* **41** 659
- [38] Thirathipviwat P, Sato S, Song G, Bednarcik J, Nielsch K, Jung J, Han J 2022 *Scr. Mater.* **210** 114470
- [39] Ma E, Wu X 2019 *Nat. Commun.* **10** 5623
- [40] Toda-Caraballo I, Rivera-Díaz-del-Castillo P E J 2015 *Acta Mater.* **85** 14

Machine learning combined with solid solution strengthening model for predicting hardness of high entropy alloys^{*}

Zhang Yi-Fan¹⁾ Ren Wei^{1)†} Wang Wei-Li^{2)‡} Ding Shu-Jian²⁾

Li Nan²⁾ Chang Liang¹⁾ Zhou Qian¹⁾

¹⁾ (School of Science, Xi'an University of Posts & Telecommunications, Xi'an 710121, China)

²⁾ (School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

(Received 21 April 2023; revised manuscript received 8 July 2023)

Abstract

Traditional material calculation methods, such as first principles and thermodynamic simulations, have accelerated the discovery of new materials. However, these methods are difficult to construct models flexibly according to various target properties. And they will consume many computational resources and the accuracy of their predictions is not so high. In the last decade, data-driven machine learning techniques have gradually been applied to materials science, which has accumulated a large quantity of theoretical and experimental data. Machine learning is able to dig out the hidden information from these data and help to predict the properties of materials. The data in this work are obtained from the published references. And several performance-oriented algorithms are selected to build a prediction model for the hardness of high entropy alloys. A high entropy alloy hardness dataset containing 19 candidate features is trained, tested, and evaluated by using an ensemble learning algorithm: a genetic algorithm is selected to filter the 19 candidate features to obtain an optimized feature set of 8 features; a two-stage feature selection approach is then combined with a traditional solid solution strengthening theory to optimize the features, three most representative feature parameters are chosen and then used to build a random forest model for hardness prediction. The prediction accuracy achieves an R^2 value of 0.9416 by using the 10-fold cross-validation method. To better understand the prediction mechanism, solid solution strengthening theory of the alloy is used to explain the hardness difference. Further, the atomic size, electronegativity and modulus mismatch features are found to have very important effects on the solid solution strengthening of high entropy alloys when genetic algorithms are used for implementing the feature selection. The machine learning algorithm and features are further used for predicting solid solution strengthening properties, resulting in an R^2 of 0.8811 by using the 10-fold cross-validation method. These screened-out parameters have good transferability for various high entropy alloy systems. In view of the poor interpretability of the random forest algorithm, the SHAP interpretable machine learning method is used to dig out the internal reasoning logic of established machine learning model and clarify the mechanism of the influence of each feature on hardness. Especially, the valence electron concentration is found to have the most significant weakening effect on the hardness of high entropy alloys.

Keywords: high entropy alloys, machine learning, genetic algorithms, solid solution strengthening

PACS: 07.05.Mh, 46.55.+d, 07.05.Kf, 61.66.Dk

DOI: 10.7498/aps.72.20230646

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 51931005, 52171048, 51571163) and the Key Industry Innovation Chain Project of Shaanxi Province, China (Grant No. 2020ZDLGY12-02).

[†] Corresponding author. E-mail: renwei@xupt.edu.cn

[‡] Corresponding author. E-mail: wlwang@nwpu.edu.cn



机器学习结合固溶强化模型预测高熵合金硬度

张逸凡 任卫 王伟丽 丁书剑 李楠 常亮 周倩

Machine learning combined with solid solution strengthening model for predicting hardness of high entropy alloys

Zhang Yi-Fan Ren Wei Wang Wei-Li Ding Shu-Jian Li Nan Chang Liang Zhou Qian

引用信息 Citation: *Acta Physica Sinica*, 72, 180701 (2023) DOI: 10.7498/aps.72.20230646

在线阅读 View online: <https://doi.org/10.7498/aps.72.20230646>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

高熵合金短程有序现象的预测及其对结构的电子、磁性、力学性质的影响

Prediction of short range order in high-entropy alloys and its effect on the electronic, magnetic and mechanical properties

物理学报. 2020, 69(4): 046102 <https://doi.org/10.7498/aps.69.20191671>

机器学习辅助绝热量子算法设计

Machine learning assisted quantum adiabatic algorithm design

物理学报. 2021, 70(14): 140306 <https://doi.org/10.7498/aps.70.20210831>

基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>

铅基钙钛矿铁电晶体高临界转变温度的机器学习研究

High critical transition temperature of lead-based perovskite ferroelectric crystals: A machine learning study

物理学报. 2019, 68(21): 210502 <https://doi.org/10.7498/aps.68.20190942>

结合机器学习的大气压介质阻挡放电数值模拟研究

Numerical study of discharge characteristics of atmospheric dielectric barrier discharges by integrating machine learning

物理学报. 2022, 71(24): 245201 <https://doi.org/10.7498/aps.71.20221555>

基于机器学习的无机磁性材料磁性基态分类与磁矩预测

Classification of magnetic ground states and prediction of magnetic moments of inorganic magnetic materials based on machine learning

物理学报. 2022, 71(6): 060202 <https://doi.org/10.7498/aps.71.20211625>