

基于人工智能的描述符预测合金催化材料形成能

李健聪¹, 王泰然², 舒武², 胡素磊¹, 欧阳润海³, 李微雪^{1,2}

(1. 中国科学技术大学合肥微尺度物质科学国家研究中心, 安徽合肥 230026;
2. 中国科学技术大学化学与材料科学学院化学物理系, 安徽合肥 230026;
3. 上海大学材料基因组工程研究院, 上海 200444)

摘要: 合金材料因其丰富可调的几何结构和电子性质, 在催化和材料科学领域得到了广泛的应用. 其中合金形成能作为一个重要的物理量, 对合金材料的形成和催化活性有重要影响. 近年来, 随着人工智能和数据库的发展, 利用机器学习的方法研究和设计新的材料成为新的研究焦点. 基于此, 通过人工智能的多任务压缩感知算法, 结合 AB₂ 合金形成能数据库展开了合金形成能描述符和预测研究. 首先建立了相应合金形成能的通用描述符, 并展开了特征敏感性分析, 揭示出合金材料组分的电子性质和几何性质的影响及其相互依赖关系. 研究结果显示, 该描述符的预测误差低于 8.10 kJ·mol⁻¹, 具有清晰的物理可阐述性, 并预测了大量未知合金材料的形成能.

关键词: 合金形成能; 数据库; 机器学习; 描述符; 敏感性分析

中图分类号: O642 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2020.06.018

引用格式: 李健聪, 王泰然, 舒武, 等. 基于人工智能的描述符预测合金催化材料形成能[J]. 中国科学技术大学学报, 2020, 50(6): 844-851.

LI Jiancong, WANG Tairan, SHU Wu, et al. AI-based descriptor for predicting alloy formation energy [J]. Journal of University of Science and Technology of China, 2020, 50(6): 844-851.

AI-based descriptor for predicting alloy formation energy

LI Jiancong¹, WANG Tairan², SHU Wu², HU Sulei¹, OUYANG Runhai³, LI Weixue^{1,2}

(1. Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China;
2. Department of Chemical Physics, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China;
3. Materials Genome Institute, Shanghai University, Shanghai 200444, China)

Abstract: Because of their rich geometric structure and electronic properties, metal alloys have been widely used in catalysis and materials science. Among them, alloys formation energy has an important influence on the formation and catalytic activity of metal alloys. With the development of artificial intelligence and databases in recent years, machine learning has been used to rationally design new materials. Based on the multi-task compressed sensing algorithm in artificial intelligence, the alloy formation energy descriptor of the AB₂ alloy formation energy database was investigated. A universal descriptor of the corresponding alloy formation energy was established, and the sensitivity analysis of features revealed the importance of electronic and geometrical properties of metal alloys. The results show that this descriptor has a prediction error lower than 8.10 kJ·mol⁻¹ and a better physical interpretation. Finally, the formation energy of a large number of unknown metal alloys was predicted.

Key words: alloy formation energy; database; machine learning; descriptor; sensitivity analysis

0 引言

在非均相催化反应中, 金属在分子的活化过程中起着重要的作用, 是目前最为常用的催化剂之

一. 对金属表面化学反应的研究与理解, 由于理论与计算化学的发展、并行计算能力的提升变得越来越高效. 相较于单金属有限的晶相、晶面实际调变空间^[1], 由多种金属组成的合金催化剂则更为多样:

收稿日期: 2020-03-27; 修回日期: 2020-05-18

基金项目: 科技部重点研发计划(2018YFA0208603), 中国科学院前沿重点项目(QYZDJ-SSW-SLH054), 国家自然科学基金重点项目(91645202, 91945302)资助.

作者简介: 李健聪, 男, 1995年生, 硕士生. 研究方向: 量子计算化学. E-mail: jcli@mail.ustc.edu.cn

通讯作者: 李微雪, 博士/教授. E-mail: wxli70@ustc.edu.cn

其组分可调^[2],不同元素的化学计量比可进行较大的调变^[3-4],具有更加丰富的晶体结构^[5-6].这些特点使得合金材料被广泛地用来优化设计相应的催化材料^[7-9],提高催化活性、选择性和稳定性,减少贵金属的用量^[10-12].

为了让合金在特定的催化反应条件下保持较优性能,研究者对合金自身的诸多物理量进行了深入的考察,其中,合金形成能作为一个与宏观机械硬度有紧密关联的物理量^[13],对于合成更稳定的合金催化剂起着重要的指导作用,同时,合金形成能说明新形成的材料是否会分相,进而指导相应材料的合成^[14].合金形成能的测定经历了“实验法测定—半经验建模估计—理论计算”的发展历程.实验方法测量合金形成能,包括直接合成法、溶液法^[15]等,通过对金属混合体系的量热,可获得在标准条件下合金的形成能.随后在 20 世纪 80 年代发展的“溶剂溶质法”^[16]在技术层面有进一步发展,能够解决部分无法直接测量的金属混合体系,但是测量的本质仍与之前的原理相同^[17].在实验测得合金形成能之后,研究者对于合金体系的热力学性质进行研究,对该体系尝试进行半经验建模^[15],其中 Miedema 等发展的模型^[18-19](平均误差为 $19.1 \text{ kJ} \cdot \text{mol}^{-1}$)^[20]最具有代表性,使用了功函以及边界电荷密度进行半经验研究,深化了对形成能问题的认识,目前,这种方法仍然是重要的形成能预测方法之一^[21-22].在理论计算方面,早期的研究者通过半经验的 EAM 方法计算合金形成能^[23-25],而随着密度泛函理论(DFT)方法的广泛应用,合金形成能够更准确地获得^[26-28],通过电子结构计算,让研究者对合金的物化性质^[29]、催化能力^[30]有更深入的理解,大大加快了合金的研发节奏^[31-32].

近年来,研究者将大量 DFT 等第一性原理程序计算的结果,在线发布于材料数据库,如 Material Project^[33]、OQMD^[34]等.由于合金研究一般以实验数据来验证结果,所以也有部分工作对合金形成能实验数据进行了汇总^[17],为进一步研究合金提供了便利.随着科学计算能力的发展,使用机器学习方法进行数据挖掘以提取有用信息和知识成为一种强有力的研究手段^[35].机器学习方法多种多样,但大多传统方法无法给出有清晰物理意义的描述符,比如常用的神经网络方法^[36-37].为了深入理解问题,寻找有意义的物理描述符的机器学习方法应运而生^[38].在催化领域,描述符是指一个能描述和预测目标性质和功能的物理量或函数,譬如 d 带中心因对分子在金属表面吸附具有线性标度^[39],因而获得了广泛应用^[40].

由于 AB_2 合金在不少电催化反应^[41-43]、储氢领域^[44-45]中性能优良,具有重要的研究价值,对合金的制备、组分调控等方面也有一定的意义,因此我们选用 AB_2 型合金的实验形成能数据进行研究.本工作展示了以真实 AB_2 组成合金为研究体系,通过能够处理异构数据的多任务学习方法获得相当准确的描述和预测合金形成能的准确预测(与真实的形成能直接比较),该描述符为二元合金形成能的

理解提供了微观认识.

1 研究方法

1.1 SISSO 方法

SISSO 是欧阳润海等最近开发的一种基于压缩感知理论的数据驱动方法^[46],可用于材料和化学等领域物理描述符的寻找.通过构建巨大特征空间,进而求解线性模型稀疏解,能够高效地找出针对目标量的最佳描述符.每个特征是基本物理量的函数,具备一定的可解释性;而最终描述符是关键特征的组合.其中,SISSO 方法的多任务学习功能^[47]可以处理多源异构数据(数据的来源多样、方法和标准不一)问题,解决因缺乏描述异构数据物理量而不能得到准确全局描述符带来的困难.目前,SISSO 方法获得了诸多成功应用,包括描述和预测无机晶体体系 Gibbs 自由形成能的物理描述符^[48]、钙钛矿材料稳定性的新容忍因子^[49]等. SISSO 方法的多任务学习策略最近被用于预测分子在金属表面的吸附^[50]以及卤化物钙钛矿带隙的预测^[51].基于数据量与物理特征的考量,本文选用多任务 SISSO 理论来研究 AB_2 合金形成能的描述符,并实现准确预测.

1.2 数据集与特征选择

我们使用了 Kim 等^[17]关于多元组成体系形成能的公开数据集,在原始的 1276 个数据点中, AB_2 组分的数据点 243 个.我们剔除重复冗余的数据点,另外,结合多任务 SISSO 算法要求对任意一类 B 金

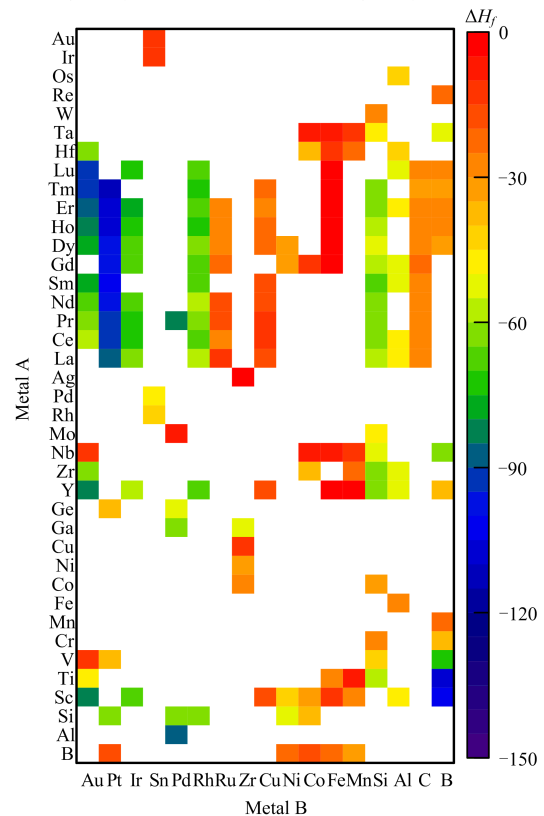


图 1 已有的 AB_2 合金形成能实验数据(单位 $\text{kJ} \cdot \text{mol}^{-1}$)
Fig. 1 Experimental formation energy data of AB_2 alloy (Unit: $\text{kJ} \cdot \text{mol}^{-1}$)

属的数据点不少于 4 个,最后得到 179 个数据(如图 1),用于下面的机器学习研究,其中 B 金属有 17 种,A 则有 39 种.

我们将清洗后的数据集随机划分为两部分:75% 的数据作为训练集,25% 的数据用于测试.关于特征集,由于 AB_2 组成的二元合金形成能与元素 A、B 同时相关,我们选用了:①几何特征:原子半径 r 、原子体积 V 、晶格常数 a ;②电子性质:Pauling 电负性 χ 、原子的第一电离能 IE 、原子核电荷数 Z 、原子极化率 α ;③热力学特征:纯金属的形成能 H_f ,其表达式为 $H_f = \Delta_f H^\circ(298.15 \text{ K, gas}) - \Delta_f H^\circ(298.15 \text{ K, crystal})$ [52-53],以下简称为 H . 上面共有 8 种基本的物理量,由于包含两种元素,所以总共有 16 种基本特征.为了描述 A、B 两种元素之间差异对所组成合金形成能的贡献,我们取同一种物理量差,构建出新的特征 ΔF ,具体包括 $\Delta F = F^B - F^A$, $F = r, V, a, \chi, IE, Z, \alpha, H$.

特征筛选是机器学习训练前的关键一步.我们采用后向筛选(backward elimination)方法对特征集进行筛选[54]:对含有 N 个特征的原始集合,构造 N 个包含 $N-1$ 个特征子集,对比子集的拟合误差,选出误差最小的最优子集;循环多次,直到误差减小到最低时,对应的子集即为最优特征集.若某一步得到的最优子集有多个(最小拟合误差完全相同),意味着删除此多个特征的效果相同,这些特征可被确定为冗余特征.

2 结果与讨论

2.1 特征集与复杂度

我们将 24 个特征作为训练集基础特征,此时训练得到的误差接近 $6.0 \text{ kJ} \cdot \text{mol}^{-1}$.我们发现去除任意一个特征之后,拟合误差会略有减少,见图 2(a)蓝线最右(右下)的数据点.这说明原先所选择的特征有冗余,冗余特征的存在会阻碍非冗余特征的优化.通过后向特征选择,我们获得的最优特征集元素个数从未经筛选的 24 个缩小到 5 个(见图 2(a)的 $N=5$ 处),拟合误差稍下降到 $5.9 \text{ kJ} \cdot \text{mol}^{-1}$.最优特征集的组成为: $\{\chi, \Delta H, \Delta V, V, \Delta Z\}$.

为了分析最优特征集里面的特征重要性,用相似的后向搜索方法对特征集缩小进行分析,我们发现按照特征的重要性进行排序,结果为 $\chi < \Delta H < \Delta V < V < \Delta Z$.加入重要性越弱的特征,对拟合误差的改进越微小,但对机器学习的泛化能力会有不小提升.譬如,对于特征集 $\{\Delta H, \Delta V, V, \Delta Z\}$,加入 A 原子的 Pauling 电负性 χ 后,描述符拟合的最大绝对误差下降近 $10 \text{ kJ} \cdot \text{mol}^{-1}$ (图 2(a)上方黑线),改善了在训练数据上的拟合能力.综上,从最优特征集进行剔除的结果可以看到,对于合金形成能的变化,影响最大的因素是两种金属电子性质的差(ΔZ 为代表),其次是与几何效应相关的因素,如金属原子 A 的体积 V 以及体积之差 ΔV .

为了确定超参数,本工作使用 k 折交叉验证[55-56] ($k=13$):每次留出一种金属的数据点作为验证集,其余的数据用来训练描述符,在得到的描

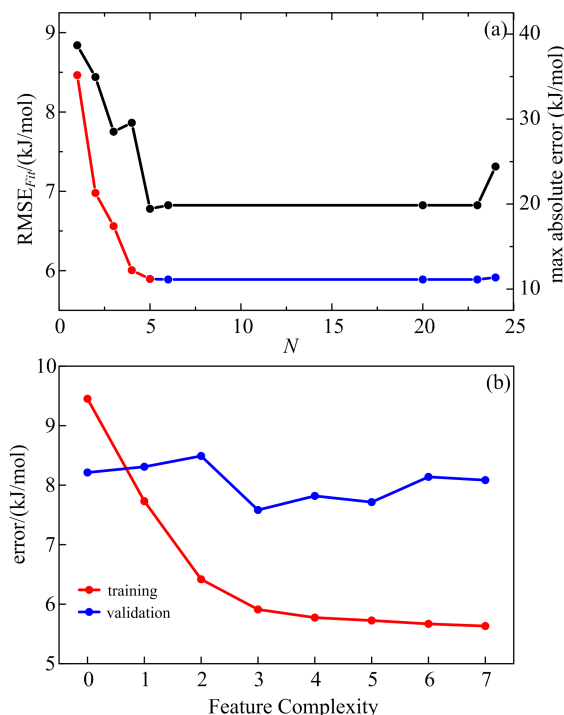


图 2 (a) 特征集的后向筛选过程中,拟合平均误差(左轴)、最大误差(黑线、右轴)变化情况;(b) 特征复杂度选择:不同特征复杂度下训练误差(红线)和验证误差(蓝线)
Fig. 2 (a) Fitting error and max absolute error of backward elimination in feature selection. (b) Feature complexity determination: training error and validation error

述符上预测数据点的值,与验证集进行比较,获得的误差取平均值.在机器学习领域中,在数据集规模相对不是很大的情况下,常采用适当提高 k 折的数量来减少对数据分布采样的偏差,实现对超参数较为合理的选择.本文训练集中 B 类金属恰好有 13 种,因此,我们选择 13 折交叉验证来确定超参数.事实上,对于多任务学习,构成描述符的关键特征越多,对应的精度可能上升,但是对描述符的理解要困难得多.由于单个关键特征的拟合误差均大于 $10 \text{ kJ} \cdot \text{mol}^{-1}$,我们选择关键特征数量为 2.本工作最核心的超参数是特征的复杂度,即一个特征中包含的数学算符个数[57].以本工作中出现的特征为例, $[(\Delta V/V)/Z]$ 的复杂度为 2, $[(\Delta V/\Delta Z) \times (\chi)^3/(V)^{-1/3}]$ 的复杂度为 5.选用不同的复杂度,特征的拟合能力也有所不同.一般来讲,复杂度越高,机器学习描述符的拟合能力也会提高,但其物理可阐释性会有所削弱.这与神经网络中的“参数量上升,拟合能力越强”类似.由于要同时考虑机器学习描述符是否存在过拟合的问题,所以需要验证误差作为指标来判断.本文发现随着复杂度的提升,描述符的拟合(训练)误差确实是在下降,然而复杂度并非越大越好.当复杂度为 3 时验证误差最小,但超过 3 时验证误差又上升(图 2(b)).因此,我们选择复杂度为 3,作为多任务 SISSO 的描述符超参数.

2.2 描述符建立

在上述得到的特征和复杂度的基础上,经过多任务 SISSO 方法,我们获得 AB_2 组成的合金形成能表达式为

$$\Delta H_f = c_1 \frac{\chi}{\Delta Z} \frac{\Delta V}{V} + c_2 \frac{\Delta H}{\chi V^3} + c_0.$$

其中, c_1, c_2, c_0 是与金属 B 有关的带有量纲的系数,具体数值见表 1. 形成能越负,形成的合金放热越多,热力学上更稳定. 另外,由于目前得到的模型本身没有包括有关有序和无序合金相关的信息,我们预期该模型不能区分相应的有序和无序合金.

表 1 AB_2 合金形成能的表达式系数

Tab. 1 Coefficients of AB_2 alloy descriptor model							
金属	Au	B	C	Co	Cu	Fe	
c_1	298.3	-1816	-406.2	77.39	-3.464	98.27	
c_2	-306.4	76.51	-103.6	-10.04	-234.7	30.17	
c_0	-67.43	-4.019	-17.21	-25.10	-19.40	-5.087	
金属	Ni	Pd	Rh	Ru	Si	Sn	Zr
c_1	110.9	-325.2	20.06	389.0	-243.2	-63.15	-277.9
c_2	-12.99	-323.6	-252.2	-71.27	-67.44	-11.65	277.3
c_0	-35.89	-70.40	-59.34	-34.96	-58.12	-19.16	-72.99

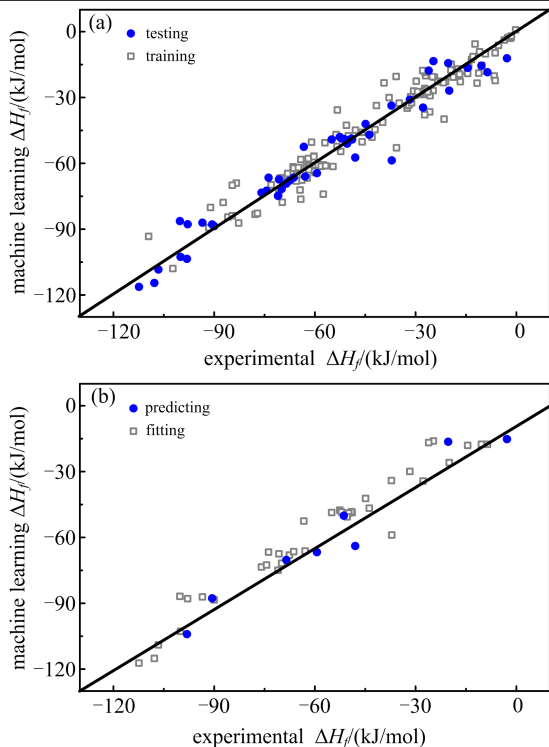


图 3 (a) 特征复杂度 = 3 时, SISSO 对 AB_2 合金形成能的训练结果(灰点)和测试结果(蓝点); (b) 测试数据中的数据用于描述符系数拟合(灰点)、预测结果(蓝点)

Fig. 3 (a) Descriptor model fitting performance in training set and testing set (Feature complexity = 3). (b) Fitting and predicting performance using testing set

该描述符在训练集上的拟合误差 $6.01 \text{ kJ} \cdot \text{mol}^{-1}$, 在测试集上的拟合误差为 $6.41 \text{ kJ} \cdot \text{mol}^{-1}$ (图 3(a)). 建立描述符既可用于全部测试数据进行拟合,也可以使用部分测试数据拟合系数后,将预

测值与真实值进行比较. 具体来说,我们留出测试集中的少量数据($\sim 20\%$),用其余的数据对描述符的系数进行较为准确的拟合,然后通过此模型直接预测留出样本的值,再与其真实实验值进行比较,如图 3(b)所示. 如此,通过预测实验得到的合金形成能数据值,可得到描述符的预测误差为 $8.10 \text{ kJ} \cdot \text{mol}^{-1}$. 这说明,经过多任务学习的描述符具有优良的预测能力;同时,此预测误差值充分说明,该描述符能够较准确地对 AB_2 组成的合金形成能的实验数据直接进行预测. 该结果比 Miedema 理论(预测误差 $19.1 \text{ kJ} \cdot \text{mol}^{-1}$)以及基于 Miedema 理论修正的统计理论(预测误差 $12.5 \text{ kJ} \cdot \text{mol}^{-1}$)^[20]能更加准确地描述合金的形成能. 另外,建立的描述符在确定超参数时,由于训练数据一共有 13 类 B 金属,在 13 折交叉验证的结果中,有 9 次的训练结果都给出了与我们最终获得的描述符完全相同的表达式,这意味我们获得的合金形成能的两个描述符组合是具有鲁棒性的,能够恰当地描述 AB_2 的形成能问题.

2.3 描述符敏感性分析和物理解释

为了对机器学习结果中各特征所起作用有直观的理解,我们在这里尝试进行敏感性分析. Xin 等^[58]曾经报道过神经网络上的敏感性分析,由于机器学习方法的通用性,我们可以直接采用此方法. 面对上文指出的多任务学习描述符系数需要拟合的问题,通过敏感性分析可直观地理解各基础特征的影响. 具体的做法是,将其他变量予以固定,“微调”某一个基础特征的值,看其对形成能的影响来获得该特征的敏感度,从而理解特征的重要性,相应的敏感度系数(normalized sensitivity coefficient, NSC)计算公式^[58]如下:

$$NSC_i^B = \frac{1}{m} \sum_{j=1}^m \left| \frac{\partial \Delta H_j}{\partial z_j^i} \right| \cdot \frac{\max(z^i) - \min(z^i)}{\max(\Delta H) - \min(\Delta H)}.$$

其中, z_j^i 是第 j 个数据样本对应的第 i 号特征, ΔH_j 是与第 j 个金属元素 A 形成 AB_2 合金的形成能, m 是金属 B 所含有的数据点个数. 对多个特征扰动同样的比例,形成能响应的变化越剧烈,则敏感度系数的绝对值越大,说明该金属 B 构成的 AB_2 合金形成能受到特征的影响越强,该特征的贡献越大. 由此,我们可以获得 AB_2 合金中对于不同的 B 元素,形成能对每个基础特征变化的敏感程度,如图 4 所示.

由图 4 可以看到,除了少数元素之外,不管是 Zr 基(敏感度系数 3.48)、Pt 基(2.48)合金,还是 IIIA、IVA 族的合金, ΔZ 的敏感性系数都比其余 4 种基础特征更大,说明形成能显著地受到电荷数差值的影响. 对于过渡金属,电荷数之差 ΔZ 、原子体积 V 至少有一个起主导影响作用. 譬如,Co 基合金对 ΔZ 、 V 的敏感度系数分别为 1.23、0.72; Ir 基则分别为 0.90(ΔZ)、1.32(V). 对于主族元素所构成的 AB_2 合金,形成能则与 A 元素的原子体积 V 敏感性较弱,如 C 基对 V 的系数为 0.71, Al 基甚至不足 0.45. 敏感度不同的原因主要是: A 金属(过渡金属为主)的原子体积基本居于主族元素原子体积的

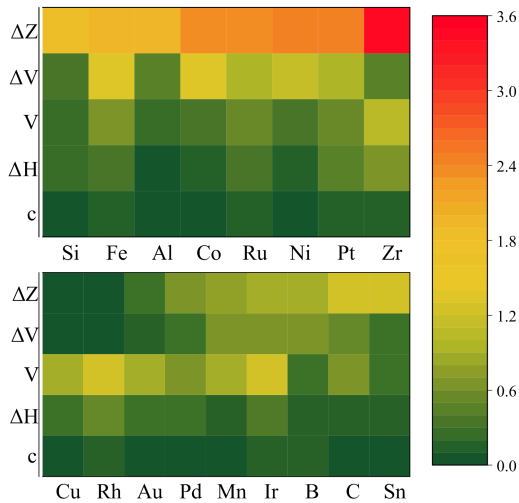


图 4 对于 AB_2 合金形成能, 五个特征在不同金属 B 类别上的敏感性系数

Fig. 4 Sensitivity coefficients of five key feature on AB_2 alloy formation, categorized by metal B

两端之间, 譬如硼的摩尔体积 4.39 cm^3 /碳 5.29 cm^3 /硅 12.06 cm^3 , 而钴 6.67 cm^3 /金 10.21 cm^3 , 导致 A 元素改变对应的几何方面贡献较难体现出来。

通过敏感性分析可知: ①在绝大多数 B 金属上, A、B 两种原子的原子核电荷数差 ΔZ 是影响形成能的最重要因素; ②而在有些情况下, A 金属原子体积 V 、原子体积之差 ΔV 至少一个在 AB_2 合金形成能也起主导作用; ③相比 ΔZ 、 V 、 ΔV , 金属 A-B 的形成能之差 ΔH 、金属 B 的电负性 χ 的敏感性相对较弱, 对能量变化的总贡献较少。

敏感性分析的作用是展示基础特征与预测值之间的依赖关系, 但是单纯通过固定其他特征, 理解某一具体特征的局部影响, 不足以反映合金形成能问题的全貌。由于 SISO 自身还是一个基于描述符的机器学习方法, 能够明确地给出描述符的表达式, 所以我们接下来对于特征所构成的描述符公式进行进一步分析。

对于 AB_2 组成的合金形成能, SISO 给出的描述符由两个关键特征构成, 我们对表达式的结构进行分析: 第一项是 $(\chi/\Delta Z)(\Delta V/V)$, 分别对应材料的电子性质和几何性质的依赖关系。由于电负性 χ 的变化范围较小(位于 $1.0 \sim 2.4$ 之间), 而其余的物理量 V (范围为 $4.6 \sim 22.5$)、 ΔZ 、 ΔV 变化较大, 特别是对于不同的 A 金属, ΔZ 、 ΔV 取值可以正负异号。这说明 V 、 ΔZ 、 ΔV 对于 AB_2 合金形成能影响最为重要, 与敏感性分析的结论①③相同。第二个关键特征 $(\Delta H)/(\chi V^3)$ 则是对第一项的修正, 其正比于 AB 两种金属的形成能之差 ΔH 即 $H^B - H^A$, 反比于 A 的电负性和体积的立方乘积。对于第二个关键特征前系数 $c_2 > 0$ 的 B 金属, 在 (χV^3) 的变化量不大时, 选用 H^A 越大的 A 金属, 合金形成过程的放热倾向更大, 说明 A 与此种 B 金属形成合金的倾向性更高; 而 $c_2 < 0$ 的情况正相反。

有了上述认识, 可尝试与其他描述符比照, 譬

如半经验的 Miedema 理论的具体公式^[58]如下:

$$\Delta H_f = \frac{2f(c)(c_A V_A^{2/3} + c_B V_B^{2/3})}{(n_{WS}^A)^{-1/3} + (n_{WS}^B)^{-1/3}} \times [-P(\Delta\phi^*)^2 + Q(\Delta n_{WS}^{1/3})^2].$$

我们可以看到, 该理论考虑了形成合金的两种金属原子的调整电负性差值 $\Delta\phi^*$ 、电子密度差值 Δn_{WS} 的作用^[59]。在此, 无论是传统方法^[26]还是本文使用的机器学习方法, 都包含了电荷以及几何方面的信息。然而, 两种描述符的差异也是明显的: ① Miedema 理论不包含 AB 两种金属的形成能之差 ΔH , 而本文的描述符中含有这一物理量, 这说明包含其他必要的物理特征(如热力学特征)在很大程度上能够弥补只使用电荷、几何相关特征的缺漏, 进一步提高预测的精确性; ②通过描述符方法建立的描述中, 电荷、几何相关的特征组合比 Miedema 描述符的表示复杂, 公式不同, 反映了若干特征需要共同作用以准确描述合金形成。

应该意识到, 通过直接分析机器学习结果中两个关键特征所起的作用, 虽可以获得对特征及其结构的描述符的进一步理解, 但是由于系数、特征的正负会导致趋势的变化难以分开理解, 而且更进一步地, V^A 、 Z^A 都是金属 A 自身的物理量, 数值不可能独立地变化。这一方面反映了合金形成能对电子效应、几何效应的依赖关系, 及形成能问题本身的复杂性; 另一方面也验证了机器学习对关键特征进行筛选的必要性。

2.4 预测合金形成能

由于数据集中的 AB_2 合金数据大多是通过直接合成法得到的, 我们可以直接对于类似方法合成的 AB_2 组成的合金形成能进行预测。由于在原始数据集中 A 金属有 39 种, B 金属 17 种, 为了保证预测的准确性, 我们在预测过程选用相同的 A、B 金属范围。局部白色的数据点代表了 AB 为同种金属的(无法形成合金, 形成能无意义)、合金形成能不在 $-150 \sim 0 \text{ kJ} \cdot \text{mol}^{-1}$ 范围内(偏离统计学习分布范围)的情况, 预测的数据如图 5 所示。

分析 B 基金属的局部预测值可以看到, 对 APt_2 和 APd_2 两种合金的合金形成能, 对金属 A 的依赖类似, 尤其是在稀土元素区域部分。值得一提的是, 稀土元素与 Pt 形成 APt_2 合金, 相关的实验数据大量缺失, 现在通过多任务 SISO 方法, 能够很好地补充了相关的数据, 而稀土合金的第一性原理计算精度相对较差。事实上, 化学计量比为 1:1 组成的 AB 合金实验上有较为完整的形成能数据, 这些数据表面相应的 Pt 基和 Pd 基的合金形成能在大范围上是一致的^[60]。类似地, 在图 5 中 ARh_2 和 AIr_2 合金形成能也具有相似的递变规律, 两种合金的形成能 Pearson 相关系数达到 0.72, 具有显著相关性。这些都共同反映了同一族元素具有相似的合金形成能力。

在 AB_2 组成的合金形成能预测图中, 通过分析总体数据分布, 还发现有如下的递变趋势: 改变金属 A 的形成能平均标准差(σ)为 $17.87 \text{ kJ} \cdot \text{mol}^{-1}$, 而固定 A 改变 B 金属对应的 $\sigma = 27.41 \text{ kJ} \cdot \text{mol}^{-1}$,

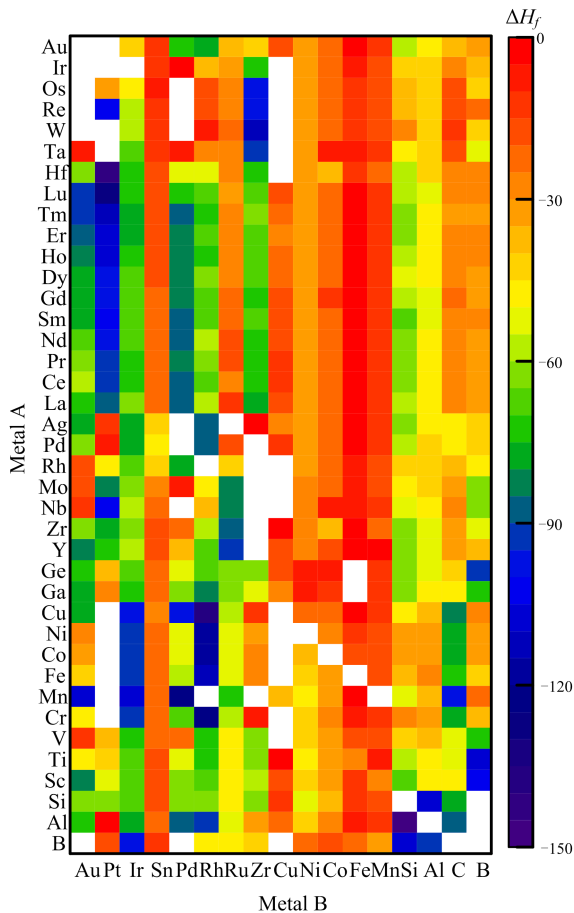


图 5 基于多任务 SISSO 得到的描述符预测的 AB_2 合金形成能(单位 $\text{kJ} \cdot \text{mol}^{-1}$)

Fig. 5 Formation energy data of AB_2 alloy based on multi-task SISSO prediction. (Unit: $\text{kJ} \cdot \text{mol}^{-1}$)

这主要是由于金属 B 含量较高所致. 当金属 B 为第三周期过渡元素时, 合金形成能的绝对值相比其他金属(主族元素、其他周期元素)低, 而且随 A 变化的波动较小. 如在图 5 中, 第三周期的 Cu 基形成能平均为 $-18.49 \text{ kJ} \cdot \text{mol}^{-1}$ ($\sigma = 7.95 \text{ kJ} \cdot \text{mol}^{-1}$)、Fe 基为 $-7.16 \text{ kJ} \cdot \text{mol}^{-1}$ ($\sigma = 6.48 \text{ kJ} \cdot \text{mol}^{-1}$), 而 C 基平均 $-42.52 \text{ kJ} \cdot \text{mol}^{-1}$ ($\sigma = 23.19 \text{ kJ} \cdot \text{mol}^{-1}$)、Pt 基平均有 $-71.71 \text{ kJ} \cdot \text{mol}^{-1}$ ($\sigma = 38.79 \text{ kJ} \cdot \text{mol}^{-1}$). 最后, 对于同属镧系元素的金属 A 而言, 与相同金属 B 形成的合金形成能也相近, 相应的 σ 仅为 $3.53 \text{ kJ} \cdot \text{mol}^{-1}$.

3 结论

本文提出了一种基于描述符学习的方法用于 AB_2 合金形成能的预测. 通过构造描述符, 可获得低于 $10 \text{ kJ} \cdot \text{mol}^{-1}$ 的预测误差, 优于传统的半经验方法及其改进描述符. 基于数据驱动描述符预测合金形成能的方法具有优良的预测能力和鲁棒性. 传统基于 Miedema 经验理论认为合金形成能可以分解到电子、几何的差异, 而基于数据驱动描述符的方法可以得知, 合金形成能的电子、几何效应是相关的. 最后我们对于 AB_2 合金形成能进行预测, 对整体理解合金形成能问题提供了基础, 为优化合金

催化剂提供更深入的认识.

参考文献(References)

- [1] MITTASCH A, FRANKENBURG W. Early studies of multicomponent catalysts [J]. *Advances in Catalysis*, 1950, 2(6): 81-104.
- [2] SINFELT J H. Catalysis by alloys and bimetallic clusters [J]. *Accounts of Chemical Research*, 1977, 10(1): 15-20.
- [3] SINFELT J H. Structure of metal catalysts [J]. *Reviews of Modern Physics*, 1979, 51(3): 569.
- [4] XU F Y, DENG S B, XU J, et al. Highly active and stable Ni-Fe bimetal prepared by ball milling for catalytic hydrodechlorination of 4-chlorophenol [J]. *Environmental Science & Technology*, 2012, 46(8): 4576-4582.
- [5] HUANG X, LI Y, et al. Synthesis of PtPd bimetal nanocrystals with controllable shape, composition, and their tunable catalytic properties [J]. *Nano Letters*, 2012, 12(8): 4265-4270.
- [6] PARAMSOTHY M, GUPTA M, SRIKANTH N. Processing, microstructure, and properties of a Mg/Al bimetal macrocomposite [J]. *Journal of Composite Materials*, 2008, 42(24): 2567-2584.
- [7] GAO W L, JIN R C, CHEN J X, et al. Titania-supported bimetallic catalysts for photocatalytic reduction of nitrate [J]. *Catalysis Today*, 2004, 90(3/4): 331-336.
- [8] FERNÁNDEZ J L, WALSH D A, BARD A J. Thermodynamic guidelines for the design of bimetallic catalysts for oxygen electroreduction and rapid screening by scanning electrochemical microscopy. M-Co (M: Pd, Ag, Au) [J]. *Journal of the American Chemical Society*, 2005, 127(1): 357-365.
- [9] STAMENKOVIC V R, FOWLER B, MUN B S, et al. Improved oxygen reduction activity on $\text{Pt}_3\text{Ni}(111)$ via increased surface site availability [J]. *Science*, 2007, 315(5811): 493-497.
- [10] CRISAFULLI C, SCIRE S, MAGGIORE R, et al. CO_2 reforming of methane over Ni-Ru and Ni-Pd bimetallic catalysts [J]. *Catalysis Letters*, 1999, 59(1): 21-26.
- [11] DE S, ZHANG J, LUQUE R, et al. Ni-based bimetallic heterogeneous catalysts for energy and environmental applications [J]. *Energy & Environmental Science*, 2016, 9(11): 3314-3347.
- [12] LEE J H, KATTEL S, JIANG Z, et al. Tuning the activity and selectivity of electroreduction of CO_2 to synthesis gas using bimetallic catalysts [J]. *Nature Communications*, 2019, 10(1): 1-8.
- [13] SCHWAB G. Some new aspects of the strength of alloys [J]. *Transactions of the Faraday Society*, 1949, 45: 385-396.
- [14] WANG C P, LIU X J, OHNUMA I, et al. Formation of immiscible alloy powders with egg-type microstructure [J]. *Science*, 2002, 297(5583): 990-993.
- [15] COLINET C. High temperature calorimetry: Recent developments [J]. *Journal of Alloys and Compounds*, 1995, 220(1/2): 76-87.
- [16] TOPOR L, KLEPPA O J. Standard molar enthalpy of formation of LaB_6 by high-temperature calorimetry [J]. *The Journal of Chemical Thermodynamics*, 1984, 16(10): 993-1002.
- [17] KIM G, MESCHEL S V, NASH P, et al. Experimental

- formation enthalpies for intermetallic phases and other inorganic compounds [J]. *Scientific Data*, 2017, 4 (1): 170162.
- [18] MIEDEMA A R, CHTEL P F D, BOER F. Cohesion in alloys — fundamentals of a semi-empirical model [J]. *Physica B+C*, 1980, 100(1): 1-28.
- [19] MIEDEMA A R, BOER F R, CHATEL P F. Empirical description of the role of electronegativity in alloy formation [J]. *Journal of Physics F: Metal Physics*, 1973, 3(8): 1558.
- [20] ZHANG R F, RAJAN K. Statistically based assessment of formation enthalpy for intermetallic compounds [J]. *Chemical Physics Letters*, 2014, 612: 177-181.
- [21] ZHANG R F, ZHANG S H, HE Z J, et al. Miedema calculator: A thermodynamic platform for predicting formation enthalpies of alloys within framework of Miedema's Theory [J]. *Computer Physics Communications*, 2016, 209: 58-69.
- [22] PALINA N, SAKATA O, KUMARA L, et al. Electronic structure evolution with composition alteration of Rh_xCu_y alloy nanoparticles [J]. *Scientific Reports*, 2017, 7(1): 1-9.
- [23] FOILES S M, BASKES M I, DAW M S. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys [J]. *Physical Review B*, 1986, 33(12): 7983.
- [24] JOHNSON R A. Alloy models with the embedded-atom method [J]. *Physical Review B*, 1989, 39 (17): 12554.
- [25] LEI Y W, SUN X R, ZHOU R L, et al. Embedded atom method potentials for Ce-Ni binary alloy [J]. *Computational Materials Science*, 2018, 150: 1-8.
- [26] WILLIAMS A R, GELATT C D, MORUZZI V L. Microscopic basis of miedema's empirical theory of transition-metal compound formation [J]. *Physical Review Letters*, 1980, 44(11): 394-395.
- [27] ZHANG Y, KRESSE G, WOLVERTON C. Nonlocal first-principles calculations in Cu-Au and other intermetallic alloys [J]. *Physical Review Letters*, 2014, 112(7): 75502.
- [28] FARSI L, DESKINS N A. First principles analysis of surface dependent segregation in bimetallic alloys [J]. *Physical Chemistry Chemical Physics*, 2019, 21(42): 23626-23637.
- [29] REITH D, PODLOUCKY R. First-principles model study of the phase stabilities of dilute Fe-Cu alloys: Role of vibrational free energy [J]. *Physical Review B*, 2009, 80(5): 54108.
- [30] OU L. The origin of enhanced electrocatalytic activity of Pt-M (M = Fe, Co, Ni, Cu, and W) alloys in PEM fuel cell cathodes: A DFT computational study [J]. *Computational and Theoretical Chemistry*, 2014, 1048: 69-76.
- [31] GUBAEV K, PODRYABINKIN E V, HART G L, et al. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials [J]. *Computational Materials Science*, 2019, 156: 148-156.
- [32] DOAK J W, HAO S, KIRKLIN S, et al. Computational prediction of nanostructured alloys with enhanced thermoelectric properties [J]. *Physical Review Materials*, 2019, 3(10): 105404.
- [33] 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): 11002.
- [34] SAAL J E, KIRKLIN S, AYKOL M, et al. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD) [J]. *JOM*, 2013, 65 (11): 1501-1509.
- [35] WU X D, ZHU X Q, WU G Q, et al. Data mining with big data [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 26(1): 97-107.
- [36] ROSCHER R, BOHN B, DUARTE M F, et al. Explainable machine learning for scientific insights and discoveries [J]. *IEEE Access*, 2020, 8: 42200-42216.
- [37] LI Z, WANG S W, CHIN W S, et al. High-throughput screening of bimetallic catalysts enabled by machine learning [J]. *Journal of Materials Chemistry A*, 2017, 5(46): 24131-24138.
- [38] GHIRINGHELLI L M, VYBIRAL J, LEVCHENKO S V, et al. Big data of materials science: Critical role of the descriptor [J]. *Physical Review Letters*, 2015, 114(10): 105503.
- [39] RUBAN A, HAMMER B, STOLTZE P, et al. Surface electronic structure and reactivity of transition and noble metals [J]. *Journal of Molecular Catalysis A: Chemical*, 1997, 115(3): 421-429.
- [40] ZHAO Z J, LIU S H, ZHA S J, et al. Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors [J]. *Nature Reviews Materials*, 2019, 4: 792-804.
- [41] CHOI K, JANG Y, CHUNG D Y, et al. A simple synthesis of urchin-like Pt-Ni bimetallic nanostructures as enhanced electrocatalysts for the oxygen reduction reaction [J]. *Chemical Communications*, 2016, 52(3): 597-600.
- [42] LI H, WEN P, LI Q, et al. Earth-abundant iron diboride (FeB_2) nanoparticles as highly active bifunctional electrocatalysts for overall water splitting [J]. *Advanced Energy Materials*, 2017, 7 (17): 1700513.
- [43] YUAN W Y, ZHAO X S, HAO W J, et al. Performance of surface-oxidized Ni_3B , Ni_2B , and NiB_2 electrocatalysts for overall water splitting [J]. *ChemElectroChem*, 2019, 6(3): 764-770.
- [44] ZALUSKA A, ZALUSKI L, STRÖM-OLSEN J O. Synergy of hydrogen sorption in ball-milled hydrides of Mg and Mg_2Ni [J]. *Journal of Alloys and Compounds*, 1999, 289(1/2): 197-206.
- [45] ZHANG L T, CAI Z L, YAO Z D, et al. A striking catalytic effect of facile synthesized $ZrMn_2$ nanoparticles on the de/rehydrogenation properties of MgH_2 [J]. *Journal of Materials Chemistry A*, 2019, 7 (10): 5626-5634.
- [46] OUYANG R H, CURTAROLO S, AHMETCIK E, et al. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates [J]. *Physical Review Materials*, 2018, 2: 083802.
- [47] OUYANG R H, AHMETCIK E, CARBOGNO C, et al. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO [J]. *Journal of Physics: Materials*, 2019, 2

- (2): 24002.
- [48] BARTEL C J, MILLICAN S L, DEML A M, et al. Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry [J]. *Nature Communications*, 2018, 9: 4168.
- [49] BARTEL C J, SUTTON C, GOLDSMITH B R, et al. New tolerance factor to predict the stability of perovskite oxides and halides [J]. *Science Advances*, 2019, 5(2): evva0693.
- [50] ANDERSEN M, LEVCHENKO S V, SCHEFFLER M, et al. Beyond scaling relations for the description of catalytic materials [J]. *ACS Catalysis*, 2019, 9(4): 2752-2759.
- [51] OUYANG R H. Exploiting ionic radii for rational design of halide perovskites [J]. *Chemistry of Materials*, 2020, 32(1): 595-604.
- [52] HAYNES W M. *CRC Handbook of Chemistry and Physics* [M]. Boca Raton, FL: CRC Press, 2014.
- [53] MEDVEDEV V A, COX J D, WAGMAN D D. *CODATA Key Values for Thermodynamics* [M]. New York: Hemisphere Publishing Corporation, 1989.
- [54] BLUM A L, LANGLEY P. Selection of relevant features and examples in machine learning [J]. *Artificial Intelligence*, 1997, 97(1/2): 245-271.
- [55] BREIMAN L, SPECTOR P. Submodel selection and evaluation in regression. The X-random case [J]. *International Statistical Review*, 1992, 60(3): 291-319.
- [56] RAO R B, FUNG G, ROSALES R. On the dangers of cross-validation. An experimental evaluation [C]// *Proceedings of the 2008 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2008: 588.
- [57] GHIRINGHELLI L M, VYBIRAL J, AHMETCIK E, et al. Learning physical descriptors for materials science by compressed sensing [J]. *New Journal of Physics*, 2017, 19(2): 23017.
- [58] MA X F, LI Z, ACHENIE L E K, et al. Machine-learning-augmented chemisorption model for CO₂ electroreduction catalyst screening [J]. *The Journal of Physical Chemistry Letters*, 2015, 6(18): 3528-3533.
- [59] 吴春峰, 李慧改, 郑少波, 等. 二元合金热力学模型—Miedema 模型 [J]. *上海金属*, 2011, 33(4): 1-5.
- [60] WATSON R E, BENNETT L H. Optimized predictions for heats of formation of transition-metal alloys II [J]. *Calphad*, 1984, 8(4): 307-321.