

低电阻率三元金合金材料的逆向设计

王向东¹, 徐鹏程¹, 卢 天¹, 刘秀娟², 陆文聪^{1,2}

(1. 上海大学 材料基因组工程研究院, 上海 200444)

(2. 上海大学理学院, 上海 200444)

摘 要: 金及其合金特别是三元金合金在电接触材料领域得到了广泛的应用。由于三元金合金组分和配比的复杂性, 如何高效地设计具有低电阻率的三元金合金电接触材料仍然是一个挑战。提出了一种低电阻率三元金合金材料的逆向设计方法, 该方法将机器学习的定性分类方法(模式识别最佳投影)与定量预测方法(XGBoost)相结合, 设计出比已有三元金合金材料电阻率更低的新材料。采用最大相关最小冗余(mRMR)结合 XGBoost 算法筛选出建模的特征变量; 利用模式识别逆投影方法设计了 3 个低电阻率三元金合金候选样本, 即 $\text{AuZr}_{1.95}\text{Cu}_{0.52}$ 、 $\text{AuZr}_{1.12}\text{Cu}_4$ 和 $\text{AuSc}_{1.86}\text{Cu}_{2.75}$, 并通过 XGBoost 模型估算了候选样本的电阻率。结果表明, 根据模式识别逆投影方法设计的样本具有较低的电阻率, 其电阻率负对数($-\lg \rho$)预报值分别为 6.718, 6.707 和 6.701, 均超过了原始数据集 $-\lg \rho$ 的最大值 6.68。该研究方法作为材料逆向设计的参考方法, 有助于实验数据的统计规律挖掘, 可以加快新材料设计。

关键词: 模式识别; 最佳投影识别; 三元金合金材料; 电阻率; 机器学习

中图分类号: TG146.3⁺1; TG132.2⁺1; TP181 **文献标识码:** A **文章编号:** 1674-3962(2021)04-0251-06

Inverse Design of Ternary Gold Alloy Materials with Low Resistivity

WANG Xiangdong¹, XU Pengcheng¹, LU Tian¹, LIU Xiujuan², LU Wencong^{1,2}

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

(2. College of Sciences, Shanghai University, Shanghai 200444, China)

Abstract: Gold and its alloys, especially ternary gold alloys, have been widely used in the field of electrical contact materials. Due to the complicated components and ratios of ternary gold alloy, how to efficiently design electrical contact materials of ternary gold alloy with low resistivity is still a challenge. In this work, new ternary gold alloy materials with lower resistivity were designed based on the inverse design method combining qualitative method (optimal projection of pattern recognition) with quantitative method (XGBoost). The critical features were screened out by using the maximum relevant minimum redundancy (mRMR) integrated with the XGBoost algorithm. Three candidate samples with lower resistivity, *i.e.*, $\text{AuZr}_{1.95}\text{Cu}_{0.52}$, $\text{AuZr}_{1.12}\text{Cu}_4$ and $\text{AuSc}_{1.86}\text{Cu}_{2.75}$ were designed by using the inverse projection of pattern recognition method developed in our laboratory, and the resistivity of the candidate samples was estimated by the XGBoost model. The results indicate that the predicted negative logarithms ($-\lg \rho$) of designed samples are 6.718, 6.707 and 6.701, respectively, exceeding the maximum value of 6.68 in the original data set. As a reference method for material inverse design, this research method is helpful for mining the statistical regularities in experimental data, and can accelerate the design of new materials.

Key words: pattern recognition; optimal projection recognition; ternary gold alloy materials; resistivity; machine learning

1 前 言

近年来, 机器学习(machine learning, ML)或数据挖掘已经成功地应用于材料科学研究中^[1-4]。例如, Xue 等^[5]报道了如何通过自适应设计加速寻找具有目标性能的新材料, 并进行了推理和全局优化, 以寻找具有极低热滞后的镍钛基形状记忆合金。何鹏程等^[6]报道了模式识别在核壳结构钴铝复合氢氧化物形貌可控合成中的应

收稿日期: 2020-10-22 修回日期: 2020-11-22

基金项目: 国家重点研究发展计划项目(2016YFB0700504);
上海市国际科技合作基金项目(18520723500)

第一作者: 王向东, 男, 1995 年生, 硕士研究生

通讯作者: 陆文聪, 男, 1964 年生, 教授, 博士生导师,
Email: wclu@shu.edu.cn

DOI: 10.7502/j.issn.1674-3962.202010008

用。由于三元金合金组分和配比的复杂性,到目前为止,找到一种简单、有效的方法来设计具有理想性能的新材料仍然是一项挑战。我们期望通过 ML 模型设计出性能更好的新材料,从而加速对新材料的研究。

金合金具有接触电阻低、导电性和导热性良好、噪音电平低以及抗有机气氛污染能力良好等优良的电学和化学性能^[7,8],故金合金在电接触材料方面具有很好的应用前景,尤其是在轻负荷、小接触压力使用条件下更能显示其优良的特性^[9-11]。电接触材料一般在电器开关

中使用,电阻率是其重要特性之一,因此研究低电阻率三元金合金具有非常重要的意义。

本工作采用的材料设计策略流程如图 1 所示。原始数据由 51 个三元金合金样本组成,利用最大相关最小冗余(mRMR)结合 XGBoost 算法筛选出建模的特征变量,然后应用作者实验室开发的模式识别逆投影方法设计了 3 个低电阻率三元金合金候选样本,最后通过 XGBoost 模型估算出候选样本的电阻率。

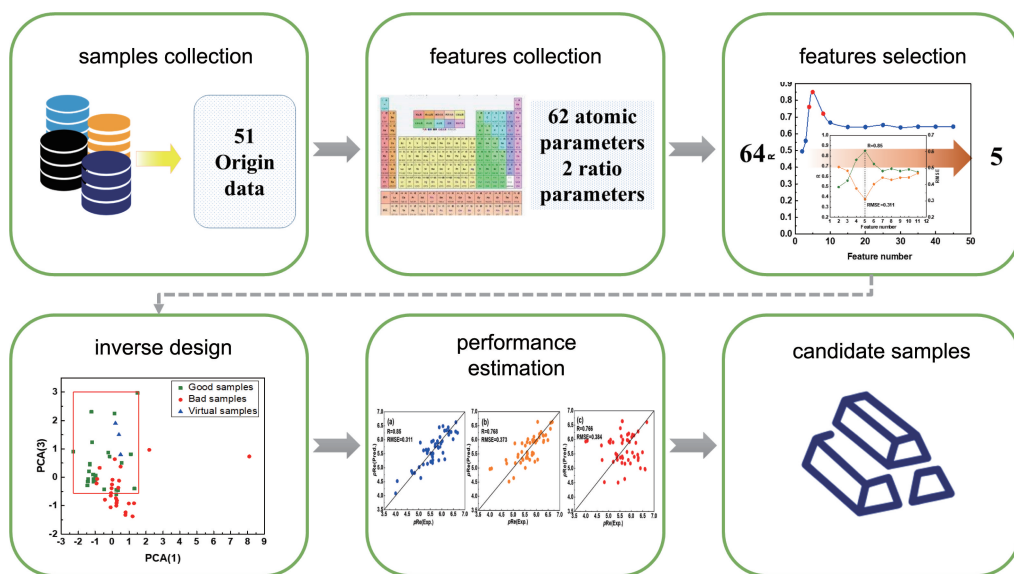


图 1 材料设计策略流程图

Fig. 1 Material design strategy flow chart

2 实验方法

2.1 数据集与描述符

本文原始数据来自于材料数据科学平台(MPDS)数据库^[12,13],它由 51 个在常温常压下的三元金合金样本组成。在数据集中,用化学符号表示三元金合金(ABC)时,先将 Au 元素排在 A 位,然后将其他两个元素按电负性上升的顺序排列,若两个元素的电负性相同,则再按价电子数上升的顺序排列。对目标值电阻率取负对数,得到其数值在 4~6.8 之间。将数据集中 $-\lg \rho$ 大于 5.71 (数据集的中位数)的划分为优类样本(good samples),将 $-\lg \rho$ 小于 5.71 的划分为劣类样本(bad samples),因此可得到优类样本 26 个,劣类样本 25 个。本工作中共收集了 64 个特征描述符^[14],其中包括 62 个原子参数描述符和 2 个组分描述符。

2.2 最大相关最小冗余算法

mRMR 算法是一种滤波式的特征筛选算法,它以不同的方式在相关性和冗余度之间进行权衡,并且以互信

息作为计算准则来比较特征与类变量之间的相关性以及特征之间的冗余度,通过最大化特征与类变量的相关性以及最小化特征之间的冗余度来进行特征选择^[15]。

最大相关性原理是指选择那些与模型具有最大相关性的特征,相关性越大,则说明训练出的模型解决问题的能力越强。特征之间的相关性越大,则冗余度越高。为了减少特征之间的冗余度并使每个特征具有代表性,需要将冗余度降低到最小,这就是最小冗余原理。

2.3 统计模式识别

统计模式识别是 ML 的主要方法之一,为了实现统计模式识别方法的自动建模,有必要从通过不同方法获得的众多投影图中自动选择最佳模式识别分类投影图(二维投影面)^[16]。

2.3.1 最佳投影识别法

由于不同的计算原理,不同的统计模式识别方法可以获得不同分类结果的投影图。但即使相同的统计模式识别方法也可能具有不同的投影方向,如主成分分析(PCA)方法能得出 $N(N-1)/2$ 个不同的投影图,其中 N

为特征变量数。为此,我们利用最佳投影识别法^[17]探索寻找分类最佳的二维投影面,其原理是在计算了若干个统计模式识别投影后(本工作应用了主成分、偏最小二乘、Fisher 判别矢量、球形映照等投影),使用迭代方法在每个隐含的投影平面上搜索出最佳分类的投影图,即在该投影图上将优类样本在一定范围内聚集,且使优化区(优类样本分布范围)混合的劣类样本的数量尽可能少。

2.3.2 逆投影法

模式识别投影图上显示的样本点的坐标是各原始特征变量的线性组合或是某种没有实际意义的映像,实际工作中实施的“优化样本”必须由原始特征变量来表示,因此需要通过特定的算法将二维模式识别图上优化区域中设计的“优化样本”返回到原始样本,该过程被称为“逆投影”^[6]。

逆投影是为二维空间的设计点找到多维空间的源像。如果没有约束条件,那么逆投影将有无数个解,故逆投影的结果只有在某些约束条件下才是唯一的。例如,为线性逆投影引入的约束条件是将设计点在各个投影矢量上的坐标取定值,而非线性逆投影引入的约束是使逆投影的误差函数最小。

本工作采用线性的模式识别逆投影方法,只需要用户在投影图上设定一个点,就能得到一组由横纵坐标的投影矢量所决定的联立方程组(含 2 个方程组),如式(1)所示:

$$\sum_{j=1}^n a_{ij}x_{ij} + b_i = c_i (i = 1, 2) \quad (1)$$

其中, x_{ij} 为第 i 个投影上的第 j 个特征变量(有 n 个), a_{ij} 和 b_i 是决定模式识别投影的系数, c_i 是设计样本的投影坐标。由式(1)确定的定量关系只有 2 个,因此,若想得到唯一解,必须给定 $n-2$ 个约束条件。本工作进一步采用 $n-2$ 个变量的平均值代入上面的方程,则可将上面的方程转化为二元一次线性方程组,从而求得该方程组的唯一解。

3 结果与讨论

3.1 变量筛选

变量筛选的目的是去除冗余的自变量,用尽可能少的自变量建立预报结果尽可能好的 ML 模型。为了去除共线性的自变量,本工作计算了所有 64 个特征描述符(自变量)之间的皮尔逊相关系数,若 2 个变量间的皮尔逊相关系数大于 0.9,则删除其中一个变量^[18]。任意 2 个描述符 x 和 y 之间的皮尔逊相关系数(R)的计算如式(2)所示:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

其中, x_i 和 y_i 分别是第 i 个样本的描述符 x 和 y 的值, \bar{x} 和 \bar{y} 分别表示这 2 个描述符的平均值。经过 R 的检验之后删除了共线性的自变量,由此得到了 44 个建模候选的描述符。

利用 mRMR 算法对 44 个特征变量进行排序^[15]。图 2 给出了排序在前 15 的变量间的皮尔逊相关系数热图。随后通过 XGBoost 算法筛选出与 ML 的最优变量子集^[19]。为了评估变量子集,采用实验值与留一法预测值之间的相关系数(R)来评价变量筛选的效果,最优变量子集相应的 R 值最大。从图 3 可以看到相关系数 R 与所选变量数之间的关系,即 R 值随着变量数先增加,在达到最大值之后逐渐减小。因为最合适的变量数可能在峰值附近,因此选择了前 11 个变量进行了更详细的计算。从图 3 中

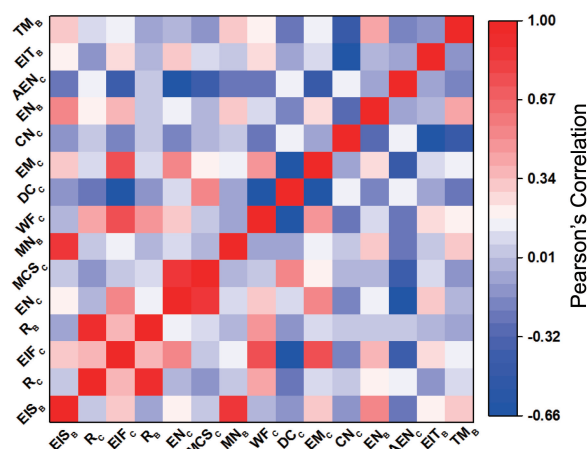


图 2 排序在前 15 的变量间的皮尔逊相关系数热图

Fig. 2 Heat map of Pearson correlation coefficient among the top 15 variables

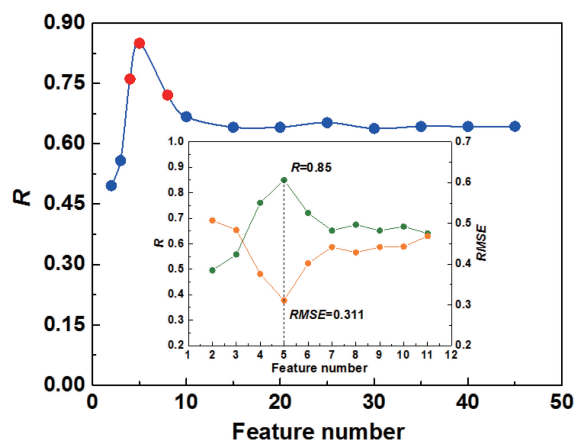


图 3 mRMR-XGBoost 进行变量筛选

Fig. 3 Variable screening by mRMR-XGBoost

可以发现, R 的趋势与均方根误差 ($RMSE$) 正好相反, 利用前 5 个变量所建的 XGBoost 模型拥有最好的表现, 即最大的 R 值与最小的 $RMSE$ 值, 故选择前 5 个变量进行后续的 ML 建模和材料设计。这 5 个变量分别为 B 位组分数(R_B)、C 位组分数(R_C)、C 位电负性(χ_C)、B 位第二电离能(I_{2B})、C 位第一电离能(I_{1C})。

3.2 逆向设计

利用本实验室的 HyperMiner 数据挖掘软件^[20], 以 R_B 、 R_C 、 I_{2B} 、 χ_C 和 I_{1C} 为特征变量, 电阻率为目标值, 经过最佳投影计算, 得到如图 4 所示的最佳模式识别投影图及逆投影点, 对应于 PCA 方法的第一主成分 PCA (1)和第三主成分 PCA(3)构成的投影图, 发现有明显的统计分布规律^[6, 21, 22]。图 4 中的矩形区域为优化区, 其中优类样本约占 70.3%, 高于总样本中优类样本所占比例(51%)。由此得出, 若要得到低电阻率的三元金合金, 则设计的三元金合金样本应尽可能控制在优化区内。图 4 中优化区分布范围可由如下联立方程组式 (3) 和式(4)表示:

4.530 ≤ 0.648[R_B] + 0.207[R_C] + 8.531 × 10⁻⁴[EIS_B] + 1.005[EN_C] + 4.358 × 10⁻³[EIF_C] ≤ 8.322 (3)

0.915 ≤ -6.979 × 10⁻³[R_B] + 7.821 × 10⁻²[R_C] + 4.678 × 10⁻³[EIS_B] - 1.775[EN_C] - 1.211 × 10⁻³[EIF_C] ≤ 4.485 (4)

在最佳模式识别投影图中选取 3 个点作为虚拟样本 (virtual samples), 如图 4 所示。然后使用模式识别逆投影的方法计算出 3 个虚拟样本的特征变量 (表 1)。最后通过计算欧式距离, 得到与虚拟样本点最接近的候选样本, 如表 2 所示。

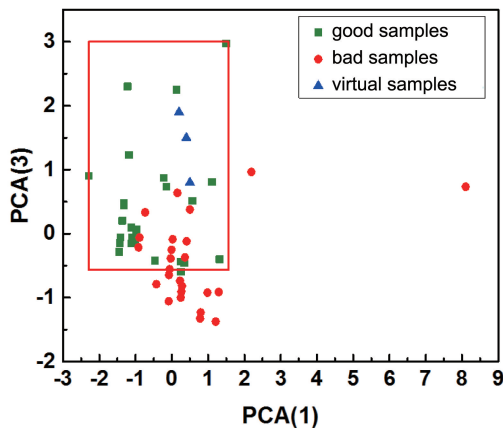


图 4 最佳模式识别投影图及逆投影点 (虚拟样本)

Fig. 4 Optimal pattern recognition projection diagram and inverse projection point (virtual samples)

表 1 逆向设计的虚拟样本

Table 1 Virtual samples of inverse design

No.	R_B	R_C	$I_{2B}/(\text{kJ} \cdot \text{mol}^{-1})$	χ_C	$I_{1C}/(\text{kJ} \cdot \text{mol}^{-1})$
1	1.95	0.52	1273.623	1.572	752.88
2	1.12	4	1278.151	1.083	751.82
3	1.86	2.75	1236.76	1.142	739.71

由表 2 可知, 候选样本分别由 1 个前过渡元素和 2 个后过渡元素组成。依据 Chen 等^[23] 研究中 3 个过渡元素金属间化合物的形成规律, 可得出结论, $R_{sp}(B)/R_{sp}(C) > 1.3$ 的三元合金系能形成三元金属间化合物 (R_{sp} 表示原子伪势半径)。因为 $R_{sp}(\text{Zr})/R_{sp}(\text{Cu})$ 为 1.38, $R_{sp}(\text{Sc})/R_{sp}(\text{Cu})$ 为 1.35, 均大于 1.3, 所以候选样本可形成三元合金。

表 2 对应虚拟样本的候选样本

Table 2 Candidate samples corresponding to virtual samples

No.	Chemical formula	R_B	R_C	$I_{2B}/(\text{kJ} \cdot \text{mol}^{-1})$	χ_C	$I_{1C}/(\text{kJ} \cdot \text{mol}^{-1})$
1	AuZr _{1.95} Cu _{0.52}	1.95	0.52	1267	1.08	745.4
2	AuZr _{1.12} Cu ₄	1.12	4	1267	1.08	745.4
3	AuSc _{1.86} Cu _{2.75}	1.86	2.75	1235	1.08	745.4

3.3 性能估算

采用 4 种不同的 ML 算法, 即 XGBoost、支持向量回归 (SVR, 采用径向基核函数)^[24]、多元线性回归 (MLR)^[25]和岭回归 (KRR)^[26], 来构建 $-\lg \rho$ 与特征变量的 ML 模型。根据每种算法的留一法交叉验证的表现筛选 $-\lg \rho$ 性能估算模型。从图 5 可以看出 XGBoost 模型拥有最高的 R 值和最低的 $RMSE$ 值, 分别为 0.850 和 0.331, 超过了其他模型的结果^[27]。因此, 后续选择 XGBoost 模型进行三元金合金 $-\lg \rho$ 的估算。

使用 XGBoost 模型对候选样本的 $-\lg \rho$ 进行估算, 可得到 3 个候选样本的 $-\lg \rho$ 值, 如表 3 所示。从表 3 中可以看出, 候选样本的 $-\lg \rho$ 值均大于原始数据集中 $-\lg \rho$ 的最大值 6.68, 故模式识别及其逆投影算法可用于低电阻率三元金合金材料的逆向设计。

表 3 候选样本估算值

Table 3 Estimated values of candidate samples

No.	Chemical formula	$-\lg \rho$
1	AuZr _{1.95} Cu _{0.52}	6.718
2	AuZr _{1.12} Cu ₄	6.707
3	AuSc _{1.86} Cu _{2.75}	6.701

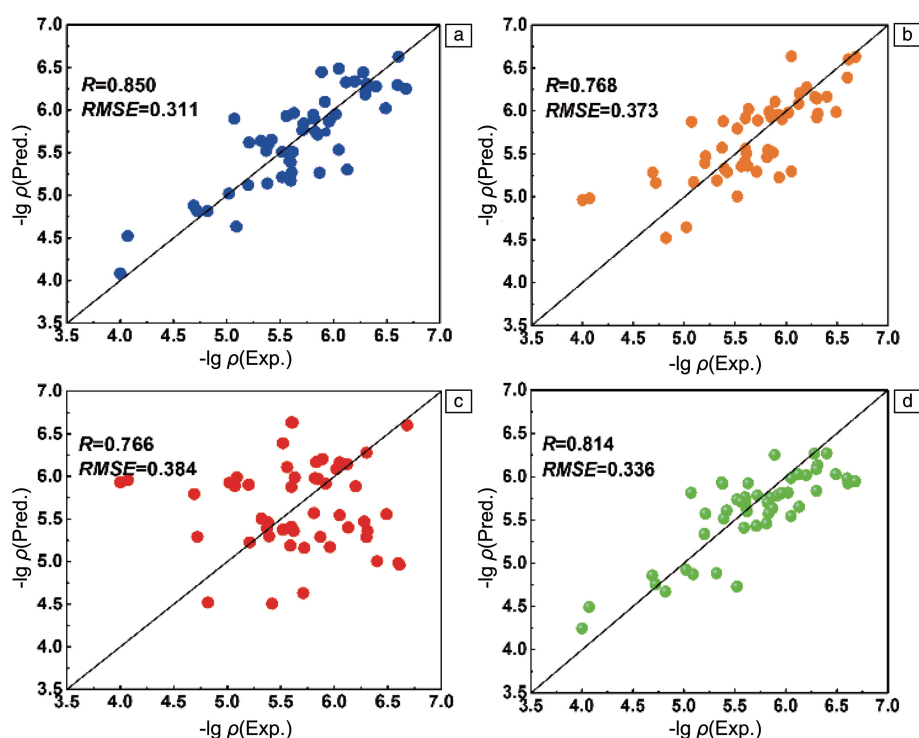


图5 采用不同 ML 算法测得的三元金合金 $-\lg \rho$ 预测值 (Pred.) 与实验值 (Exp.): (a) XGBoost, (b) KRR, (c) MLR, (d) SVR

Fig. 5 Predicted value (Pred.) and experimental value (Exp.) of $-\lg \rho$ of ternary gold alloys by using different ML algorithms: (a)

XGBoost, (b) KRR, (c) MLR, (d) SVR

4 结论

本文以设计低电阻率三元金合金为目标, 利用本实验室的 HyperMiner 数据挖掘软件, 通过模式识别最佳投影找出了形成低电阻率三元金合金的边界条件, 然后应用模式识别逆投影方法设计了 3 个低电阻率三元金合金候选样本, 最后通过 XGBoost 模型估算出候选样本的电阻率。结果表明, 根据逆投影方法设计的 $\text{AuZr}_{1.95}\text{Cu}_{0.52}$ 、 $\text{AuZr}_{1.12}\text{Cu}_4$ 和 $\text{AuSc}_{1.86}\text{Cu}_{2.75}$ 样本具有较低的电阻率, 其 $-\lg \rho$ 预报值分别为 6.718, 6.707 和 6.701, 均超过了原始数据集 $-\lg \rho$ 的最大值 6.68。因此, 本工作的研究方法可用于指导新材料的理论设计, 有助于实验数据的统计规律挖掘, 用以加快新材料设计发展。

参考文献 References

- [1] ESTEVA A, KUPREL B, NOVOA R A, *et al.* Nature[J], 2017, 542(7639): 115–118.
- [2] WEN C, ZHANG Y, WANG C, *et al.* Acta Materialia[J], 2019, 170: 109–117.
- [3] 王卓, 王礞, 雍歧龙, 等. 中国材料进展[J], 2017, 36(02): 132–140.
WANG Z, WANG M, YONG Q L, *et al.* Materials China[J], 2017, 36(02): 132–140.
- [4] 温晓灿, 张凡, 雷智锋, 等. 中国材料进展[J], 2019, 38(03): 242–250.
WANG X C, ZHANG F, LEI Z F, *et al.* Materials China[J], 2019, 38(03): 242–250.
- [5] XUE D, BALACHANDRAN P V, HOGDEN J, *et al.* Nature Communications[J], 2016, 7(1): 11241.
- [6] 何鹏程, 张庆, 陆文聪. 计算机与应用化学[J], 2016, 33(06): 642–644.
HE P C, ZHANG Q, LU W C. Computers and Applied Chemistry[J], 2016, 33(06): 642–644.
- [7] 李银娥, 姜婷, 马光, 等. 稀有金属快报[J], 2008, 27(12): 1–6.
LI Y E, JIANG T, MA G, *et al.* Rare Metals Letters[J], 2008, 27(12): 1–6.
- [8] 宁远涛. 贵金属[J], 2007(02): 57–64.
NING Y T. Precious Metals[J], 2007(02): 57–64.
- [9] 陈南光. 世界有色金属[J], 2017(15): 235–237.
CHENG N G. World Non-Ferrous Metals[J], 2017(15): 235–237.
- [10] 马晓东, 余建军, 赵涛. 山东工业技术[J], 2016(19): 255–256.
MA X D, YU J J, ZHAO T. Shandong Industrial Technology[J], 2016(19): 255–256.
- [11] 朱文哲. 中国科技信息[J], 2019(16): 68–69.
ZHU W Z. China Science and Technology Information[J], 2019(16): 68–69.
- [12] BLOKHIN E, VILLARS P. Handbook of Materials Modeling[M].

- Berlin: Springer-Verlag, 2019: 1–25.
- [13] 杨丽, 苏航, 柴锋, 等. 中国材料进展[J], 2019, 38(07): 672–681.
- YANG L, SU H, CHAI F, *et al.* Materials China[J], 2019, 38(07): 672–681.
- [14] VILLARS P, CENZUAL K, GLADYSHEVSKII R, *et al.* Materials Informatics: Methods, Tools and Applications[M]. Berlin: Springer-Verlag, 2019: 55–106.
- [15] 毕青松, 梁雪春, 陈舒期. 计算机应用与软件[J], 2020, 37(09): 296–301.
- BI Q S, LIANG X C, CHEN S Q. Computer Applications and Software[J], 2020, 37(09): 296–301.
- [16] 叶晨洲, 杨杰, 姚莉秀, 等. 科学通报[J], 2001, 46(23): 1951–1953.
- YE C Z, YANG J, YAO L X, *et al.* Chinese Science Bulletin[J], 2001, 46(23): 1951–1953.
- [17] 徐潇, 卢凯亮, 蒋若宁, 等. 上海大学学报(自然科学版)[J], 2019, 25(01): 95–100.
- XU X, LU K L, JIANG R N, *et al.* Journal of Shanghai University (Natural Science Edition)[J], 2019, 25(01): 95–100.
- [18] YUAN R, LIU Z, BALACHANDRAN P V, *et al.* Advanced Materials[J], 2018, 30(7): 1702884.
- [19] 叶倩怡, 饶泓, 姬名书. 南昌大学学报(理科版)[J], 2017, 41(03): 275–281.
- YE Q Y, RAO H, JI M S. Journal of Nanchang University (Natural Science)[J], 2017, 41(03): 275–281.
- [20] 陆文聪. HyperMiner 数据挖掘软件[EB/OL]. (2011–01–16) [2020–03–26]. <http://materials-data-mining.com/home#>.
- LU W C. HyperMiner Data Mining Software[EB/OL]. (2011–01–16) [2020–03–26]. <http://materials-data-mining.com/home#>.
- [21] 徐潇, 赵洪涛, 卢凯亮, 等. 计算机与应用化学[J], 2018, 35(06): 433–439.
- XU X, ZHAO H T, LU K L, *et al.* Computers and Applied Chemistry[J], 2018, 35(06): 433–439.
- [22] LU W C, XIAO R, YANG J, *et al.* Journal of Materiomics[J], 2017, 3(3): 191–201.
- [23] CHEN N, LU W C, QIN P, *et al.* Journal of Alloys and Compounds[J], 1999, 289: 126–130.
- [24] NAZEMI M, HEIDARIPANAH A. Road Materials and Pavement Design[J], 2016, 17(3): 768–778.
- [25] SHAFIQ M, AMIR W M, ZAFAKALI N S. Journal of Modern Applied Statistical Methods[J], 2017, 16(2): 490–505.
- [26] 罗良文, 阚大学. 世界经济研究[J], 2011(04): 31–35.
- LUO L W, KAN D X. World Economy Study[J], 2011(04): 31–35.
- [27] SHI L, CHANG D, JI X, *et al.* Journal of Chemical Information and Modeling[J], 2018, 58(12): 2420–2427.

(本文为本刊约稿, 编辑 费蒙飞)