

# 基于不同机器学习算法的铝合金性能预测

李 婷

(沈阳东创贵金属材料有限公司, 辽宁 沈阳 110000)

**[摘 要]** 铝合金由于其高强度、低重量的特点被广泛的应用于航空航天以及交通等领域,高性能铝合金的设计是当下的热点。本文以铝合金组分为输入向量,抗拉强度为目标变量,建立了 RF、ET、Bagging、Adaboost 四种不同的机器学习算法模型。结果表明:RF 模型具有最佳的预测性能, $R=0.89$ 、 $MAE=40.33$ ;Ti 元素含量对铝合金抗拉强度的预测起正向作用,Ti 元素含量越高,抗拉强度值越大;Mg 元素、Cu 元素含量对铝合金抗拉强度的预测作用并不明显;Zn 元素、Ce 元素、Y 元素含量对铝合金抗拉强度的预测起负向作用,即元素含量越大,抗拉强度值越小,特征重要性从大到小分别为  $Ti > Mg > Cu > Zn > Ce > Y$ 。

**[关键词]** 机器学习; 铝合金; 抗拉强度; 性能预测

**[中图分类号]** TF821

**[文献标志码]** B

**[文章编号]** 1003-8884(2023)04-0066-06

**DOI:** 10.19611/j.cnki.cn11-2919/tg.2023.04.013

## 0 引言

可持续的绿色发展方式是当今世界的主旋律,对材料的性能也提出了高强度、低重量的高标准要求,尤其是在航空航天领域,降低材料的重量可以减少油耗、增加航程,7xxx 系铝合金正是这样一种材料<sup>[1]</sup>。7xxx 系铝合金最早在 1932 年由维贝尔发现,随后通过调整合金成分、添加微量元素,又发展出很多新合金,使得 7xxx 系铝合金种类增多,作为高强度铝合金的代表,以其较高的强度以及良好的韧性和耐腐蚀性等优异性能而被广泛地应用于航空航天及交通运输领域<sup>[2-5]</sup>。

实验试错法是材料科学中的常用方法,即首先凭借科研工作者的个人经验先对材料的成分、结构、性能等进行设计,然后采用实验手段制备出所设计的材料,再测试其性能,从宏观维度揭示其是否具有应用价值,之后采用表征手段显示其微观形貌特点,解释其内在机理。如果设计的材料不符合预期的特点,则对其进行调整,继续进行上述流程。此种方法需要大量的人力和物力、需要消耗大量的能源、研发

周期很长、存在较大的盲目性。随着计算机算力的提升,基于第一性原理的密度泛函理论计算作为材料科学的“第三范式”,被广泛应用于新材料的开发,具有较高的准确性,但此种方法计算花费大<sup>[6]</sup>。随着人工智能技术的发展,机器学习作为材料科学的“第四范式”,成为材料研发过程的热点,为材料性能的预测<sup>[7-9]</sup>提供了见解。

铝合金的成分是影响其力学性能的关键因素,Mg、Zn、Cu 是 7xxx 系铝合金的主要添加元素,通过影响合金的微观组织变化影响其各项性能<sup>[10]</sup>。目前商用铝合金的极限抗拉强度均在 700 MPa 以下,为了开发工业化高强度 7xxx 铝合金,优化合金成分是一种实用的策略<sup>[11]</sup>。但由于铝合金组成较为复杂,不同组分的铝合金的数量变得难以估量,因此采用传统实验手段进行铝合金的设计变得十分困难。

本文收集了铝合金的组分及抗拉强度值形成初始数据集<sup>[11]</sup>,以铝合金的组分作为输入向量、抗拉强度作为目标向量,采用 RF、ET、Bagging、Adaboost 建立不同的机器学习算法模型,揭示了铝合金组分性能之间的关系,为铝合金的设计提供理论指导。

## 1 机器学习算法

### 1.1 Bagging 算法

Bagging 算法是自举汇聚法 (bootstrap aggregating) 的简写,是一种并行式的集成学习方法<sup>[12]</sup>,是

**[收稿日期]** 2023-05-15

**[第一作者]** 李婷(1990—),女,陕西安康人,中级工程师,主要从事金属材料工艺研发和设计工作。

**[引用格式]** 李婷.基于不同机器学习算法的铝合金性能预测[J].有色设备,2023,37(4):66-71.

在自助采样法(Bootstrap Sample)的基础上构建的。自助采样法是指采用随机的、有放回的取样方式,给定  $m$  个样本的数据集,随机取出一个样本放入采样集,再把样本放回至初始数据集,重复此过程  $m$  轮,则可以得到一个具有  $m$  个样本的采样集。重复  $T$  次将会产生  $T$  个采样集,然后基于每个采样集各训练出一个学习器,最后将这些学习器进行组合从而得到更好的预测结果。在预测输出时,对于分类问题采用简单投票法,对于回归问题采用简单平均法。

## 1.2 随机森林(RF)

随机森林(Random Forest, RF)是一种以决策树为基学习器的集成学习算法<sup>[13]</sup>,它既可以用于分类问题,又可以用于回归问题。基本思想为将多个弱学习器的结果进行组合,从而形成一个预测性能更好的学习器,即遵循“少数服从多数”的原则。它是 Bagging 算法的一种典型的机器学习算法模型。在树的生长过程采用有放回的抽样方式以及采用特征的最佳分割点进行划分,即采用数据随机和特征随机的原则,可以降低学习器的方差,降低模型过拟合的风险。具有鲁棒性强、并行简单、训练速度快等优点。

具体过程如下所述。

1) 输入样本集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , 对于  $t = 1, 2, \dots, T$ , 对训练集进行第  $t$  次随机采样,共采集  $m$  次得到包含  $m$  个样本的采样集  $D_T$ 。

2) 用采样集  $D_T$  训练第  $T$  个决策树模型  $G_T(x)$ 。在结点分裂形成叶子结点时,从所有特征选取一部分特征,对于分类问题,采用最大信息增益、最大基尼增益等方式选取一个最优特征进行决策树左右子树的划分,对于回归问题,采用最小平均绝对误差(MAE),最小均方误差(MSE)的方式进行左右子树的划分。

3) 将  $T$  个决策树模型组合得到最后的结果。

## 1.3 极端随机树(ET)

极端随机树(Extremely randomized trees, ET)也是一种集成学习算法<sup>[14]</sup>,既可以用于分类问题,也可以用于回归问题。与随机森林算法类似,都是由许多棵决策树构成的,但是极端随机树并不采取自助采样策略,直接使用原始训练样本,从而可以减少偏差。在每棵决策树的结点分裂时,阈值的选取是随机的。对于回归问题,当样本的特征值大于阈值时分到左分支,小于阈值时,分到右分支,然后计算

此时的均方误差或者平均绝对误差。遍历所有的特征值,得到所有特征的均方误差或平均绝对误差。根据误差最大的原则实现对结点的分裂。

## 1.4 Adaboost 算法

Adaboost 是自适应增强(Adaptive Boosting)的简写,是一种串行式的集成学习算法<sup>[15]</sup>,既可以用于分类问题,又可以用于回归问题。它是由 Freund 和 Schapire 在 1995 年提出的<sup>[15]</sup>,其核心思想为提升(boosting)思想,即通过原始数据集生成弱学习器,然后采用迭代的方式更新样本权重生成不同的弱学习器,最后采用结合策略把弱学习器结合起来生成一个强学习器。

具体过程如下所述。

1) 初始化样本权重计算式见式(1)。

$$\omega_{1i} = \frac{1}{m} \quad (1)$$

式中: $m$  为样本的数量; $\omega_{1i}$  表示第  $i$  个样本的权重; $i = 1, 2, \dots, m$ 。

2) 采用具有权重的样本集进行训练得到弱学习器  $G_k(x)$ ,  $k = 1, 2, \dots, K$ 。

计算弱学习器的回归误差率计算式见式(2)。

$$e_k = \sum_{i=1}^m \omega_{ki} e_{ki} \quad (2)$$

式中: $e_k$  为弱学习器的回归误差率; $m$  为样本个数; $\omega_{ki}$  为第  $k$  个弱学习器、第  $i$  个点的权重; $e_{ki}$  为第  $k$  个弱学习器、第  $i$  个点的相对误差。

相对误差主要有线性误差、平方误差与指数误差 3 种,分别如式(3)、(4)、(5)所示。

$$e_{ki} = \frac{|y_i - G_k(x_i)|}{E_k} \quad (3)$$

$$e_{ki} = \frac{(y_i - G_k(x_i))^2}{E_k^2} \quad (4)$$

$$e_{ki} = 1 - \exp\left(-\frac{|y_i - G_k(x_i)|}{E_k}\right) \quad (5)$$

式中: $x_i, y_i$  分别为第  $i$  个样本点的特征值与目标值。

$E_k$  为最大绝对误差,如式(6)所示。

$$E_k = \max_i |y_i - G_k(x_i)| \quad (6)$$

计算弱学习器的权重表达式见式(7)。

$$\alpha_k = \frac{e_k}{1 - e_k} \quad (7)$$

式中: $\alpha_k$  为弱学习器的权重; $e_k$  为弱学习器的回归误差率。

4) 下一次迭代的权重更新表达式见式(8)~(9)。

$$\omega_{k+1,i} = \frac{\omega_{ki} \alpha_k^{1-e_{ki}}}{Z_k} \quad (8)$$

$$Z_k = \sum_{i=1}^m \omega_{ki} \alpha_k^{1-e_{ki}} \quad (9)$$

式中: $Z_k$ 是规范化因子,使样本权重和为1; $\omega_{ki}$ 为第  $k$  个弱学习器、第  $i$  个点的权重; $\omega_{k+1,i}$ 为第  $k+1$  个弱学习器、第  $i$  个点的权重; $\alpha_k$ 为弱学习器的权重。

5) 采用结合策略,构建最终学习器,表达式见式(10)。

$$f(x) = \sum_{k=1}^K \left( \ln \frac{1}{\alpha_k} \right) g_k(x) \quad (10)$$

式中: $f(x)$ 为模型预测值; $g(x)$ 是所有  $\alpha_k G_k(x)$  的中位数; $\alpha_k$ 为弱学习器的权重; $k=1,2,\dots,K$ 。

### 1.5 性能评价指标

不同机器学习算法性能的好坏用平均绝对误差(MAE)和皮尔逊相关系数(R)来评价,分别见式(11)、式(12)。MAE 的值越小,R 的绝对值越接近于 1,表明模型的性能越好。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i| \quad (11)$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y}_i) \sum_{i=1}^n (p_i - \bar{p}_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (p_i - \bar{p}_i)^2}} \quad (12)$$

式中: $y_i$ 表示铝合金抗拉强度的实验值; $p_i$ 表示机器学习算法的预测值; $\bar{y}_i$ 表示铝合金抗拉强度实验值的平均值; $\bar{p}_i$ 表示机器学习算法预测值的平均值。

## 2 算法建立

材料的结构往往决定着性质,因此材料的组成可能与材料的某种性质存在着线性或者非线性关系。更复杂的材料组成使得其种类成倍增加,因此依靠传统的实验手段获得这些关系变得更加不可能。借助于机器学习算法则使得这个过程变得更加简单,可以揭示这种组分-结构-性质之间的关系。本文所使用的铝合金抗拉强度数据集来源于 Li 等人<sup>[11]</sup>的研究,为实验获得的真实值。以铝合金的 6 种元素(Zn, Mg, Cu, Y, Ce, Ti)组成作为输入向量,铝合金的抗拉强度作为目标向量,建立不同的机器学习算法模型,基于机器学习的铝合金抗拉强度预测模型构建方法如下所述。

1) 数据准备:将 73 条数据随机分成两部分,其中以 85% 的数据作为训练集,用于模型的训练,剩余 15% 的数据作为测试集,用于评估模型的泛化能力。

2) 模型训练:以铝合金元素组成为输入向量,抗拉强度为目标向量,建立 Bagging、RF、ET、Adaboost 算法模型,其中 Bagging 与 Adaboost 均采用决策树(Decision Tree, DT)作为基估计器。

3) 模型性能评估:采用平均绝对误差(MAE)、皮尔逊相关系数(R)对模型的性能进行评估。

4) 模型解释:采用基于博弈论的 SHAP 方法对最好的机器学习算法模型进行解释,探究各个特征对模型性能的影响。

5) 模型应用:随机生成不同组分的铝合金,采用最好的模型预测其抗拉强度。

## 3 结果与讨论

### 3.1 特征相关性分析

特征与特征之间存在高度的相关性时,会降低模型的稳定性和预测性能,特征与目标向量存在高度的相关性时,则有利于提高模型的稳定性和预测性能。因此首先需要分析特征与特征、特征与目标向量之间的相关性。本文通过特征与特征之间的皮尔逊相关系数热图和 mRMR 分数来揭示相关性规律。特征与特征之间的皮尔逊相关系数热图如图 1(左)所示,mRMR 分数如图 1(右)所示。皮尔逊相关系数是一种衡量相关性的量度,其取值在 -1 到 1 之间,绝对值越接近于 1,说明两个特征之间的相关性越高,即存在特征冗余;绝对值越接近 0,说明两个特征之间的相关性越低。mRMR(Max Relevance Min Redundancy)<sup>[16]</sup>,即最大相关最小冗余算法,可以通过迭代的方式寻找特征之间相关性最小、与目标向量相关性最大的前  $k$  个特征。从图中可以看出,特征之间的相关性很小,mRMR 分数由大到小分别为 Ti > Y > Ce > Cu > Mg > Zn,因此特征之间不存在冗余,所以可以直接用于铝合金抗拉强度的预测。

### 3.2 不同算法模型预测结果

为了选择出对铝合金抗拉强度预测性能最好的模型,本文采用 RF、ET、Bagging、Adaboost 四种机器学习算法建立了铝合金抗拉强度的预测模型,采用皮尔逊相关系数(R)、平均绝对误差(MAE)用于评

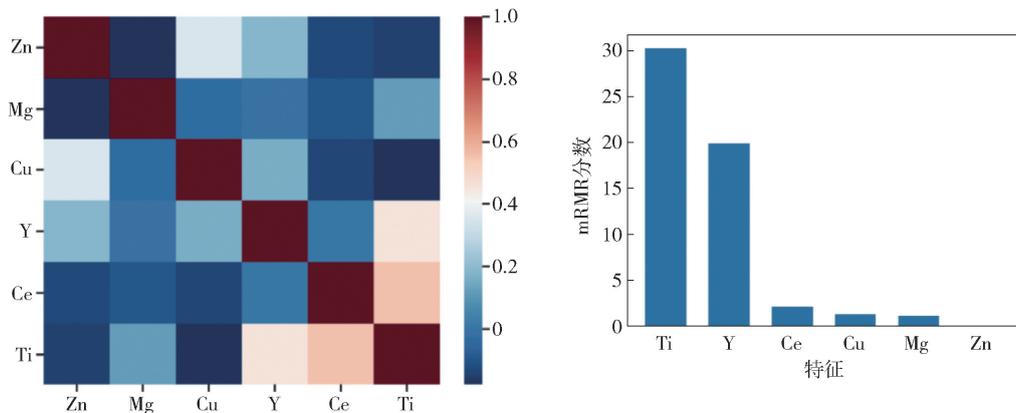


图 1 特征相关系数热图(左)及 mRMR 分数排名(右)

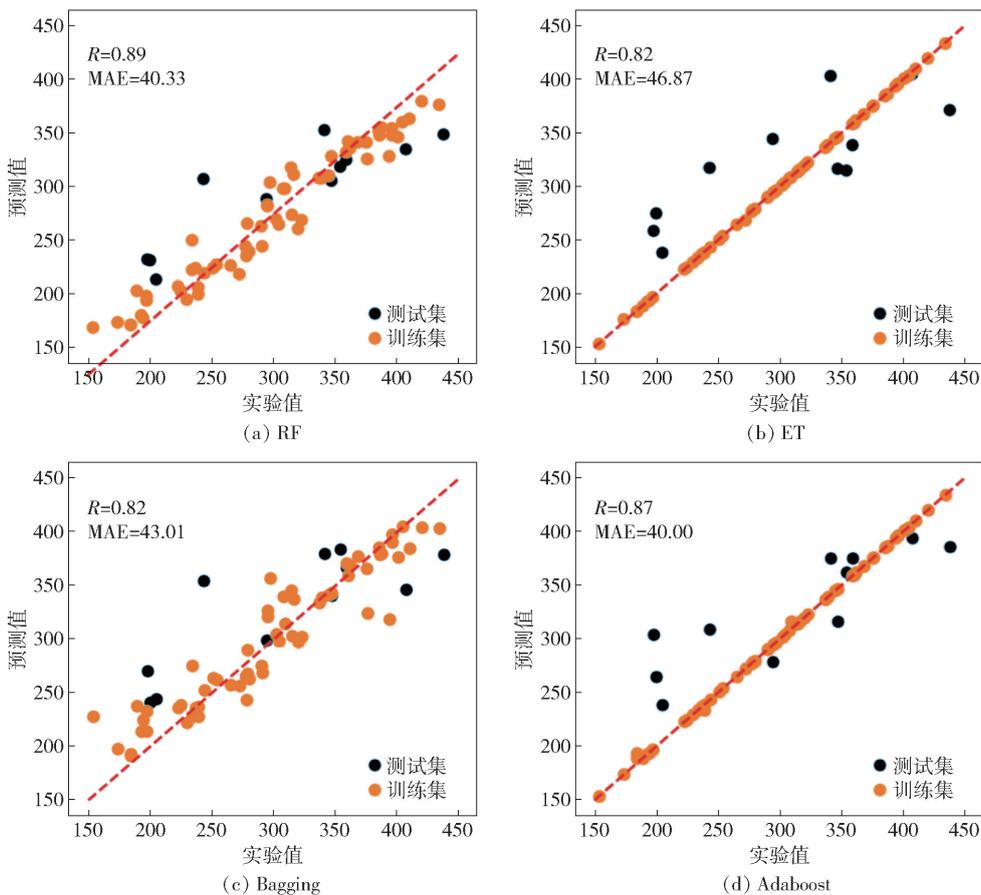


图 2 不同算法模型预测结果

价模型的性能。皮尔逊相关系数越接近于 1, 平均绝对误差越小, 说明模型的性能越好。图 2 显示了不同机器学习算法在训练集与测试集上的表现, 它们的皮尔逊相关系数均在 0.8 以上, 4 种机器学习算法的预测值与真实值均在  $y=x$  附近, 说明它们都具有很好的预测效果, 其中 RF 具有最高的皮尔逊相关系数, Adaboost 算法具有最小的平均绝对误差,

考虑到 RF 的平均绝对误差与 Adaboost 相差不大, 因此选择 RF 算法用于铝合金抗拉强度的预测。

### 3.3 模型解释

类似于 RF、ET、Bagging、Adaboost 等一些非线性的模型尽管在材料性能预测上具有很高的精度, 特征的维度相对较高(大于 3 维)时, 使得模型的解释性下降。模型解释在材料性能预测方面也十

分重要,可以使我们探索其中的机理,为新材料的开发提供见解。SHAP(Sharpley Additive explanation)是解决模型可解释性的一种方法,它是 Lundberg 和 Lee 在 2016 年提出的<sup>[17]</sup>。SHAP 基于 Shapley 值,该值是经济学家 Lloyd Shapley 提出的博弈论概念。该方法为通过计算在合作中个体的贡献来确定该个体的重要程度。图 3 是用于解释 RF 模型 SHAP 值分布的散点图,我们可以了解每个特征对模型的预测是正向的还是负向的,图中的每一个点代表了每个样本的每个特征对模型输出的贡献,颜色代表了特征值的大小的,红色越深表明特征值越大,蓝色越深表明特征值越小。从图中可以看出,Ti 元素含量对铝合金抗拉强度的预测起正向作用,Ti 元素含量越高,抗拉强度值越大;Mg 元素、Cu 元素含量铝合金抗拉强度的预测作用并不明显;Zn 元素、Ce 元素、Y 元素含量对铝合金抗拉强度的预测起负向作用,即元素含量越大,抗拉强度值越小。图 4 显示了 RF 模型的 SHAP 重要性排序,可以看出,特征重要性从大到小分别为 Ti > Mg > Cu > Zn > Ce > Y,Ti 元素含量的特征重要性最大,远远大于其他 5 种元素含量,Mg 元素、Cu 元素、Zn 元素含量的特征重要性较为接近,Ce 元素、Y 元素含量的特征重要性较小。

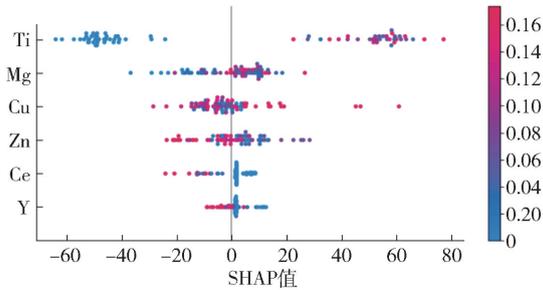


图 3 RF 模型 SHAP 值分布散点图

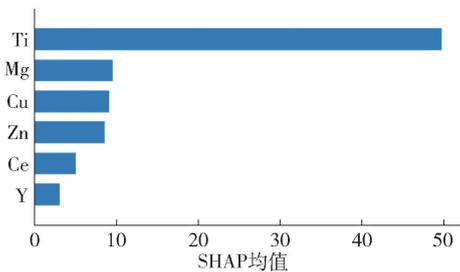


图 4 RF 模型的 SHAP 重要性

### 3.4 模型应用

在采用铝合金 6 种元素组成用于其抗拉强度的

预测时,RF 模型拥有最佳的预测效果,它的皮尔逊相关系数为 0.89,平均绝对误差为 40.33。因此可以随机生成一些虚拟样本,然后通过 RF 模型预测其抗拉强度,便于后续通过实验手段进行性能测试,继而为铝合金材料的设计提供见解。表 1 显示了原始数据集各元素含量的统计特性,我们可以在各元素含量的最小值、最大值之间生成一些虚拟样本,并采用 RF 模型对其抗拉强度进行预测,表 2 显示了一些虚拟样本的抗拉强度值。

表 1 原始数据集各元素含量的统计特性

元素	Zn	Mg	Cu	Y	Ce	Ti
最小值	5.1	1.3	1.2	0	0	0
最大值	8.3	2.9	2.5	0.49	0.48	0.19

表 2 虚拟样本的抗拉强度值

序号	Zn	Mg	Cu	Y	Ce	Ti	预测值
1	5.00	1.00	1.00	0.00	0.00	0.00	246.83
2	5.21	1.11	1.11	0.26	0.26	0.11	291.88
3	5.42	1.21	1.21	0.53	0.53	0.21	267.98
4	5.63	1.32	1.32	0.79	0.79	0.32	267.98
5	5.84	1.42	1.42	1.05	1.05	0.42	268.59
6	6.05	1.53	1.53	1.32	1.32	0.53	287.84
7	6.26	1.63	1.63	1.58	1.58	0.63	322.16
8	6.47	1.74	1.74	1.84	1.84	0.74	320.72
9	6.68	1.84	1.84	2.11	2.11	0.84	319.98
10	6.89	1.95	1.95	2.37	2.37	0.95	316.76
11	7.11	2.05	2.05	2.63	2.63	1.05	317.02
12	7.32	2.16	2.16	2.89	2.89	1.16	318.24
13	7.53	2.26	2.26	3.16	3.16	1.26	317.97
14	7.74	2.37	2.37	3.42	3.42	1.37	323.36
15	7.95	2.47	2.47	3.68	3.68	1.47	316.75
16	8.16	2.58	2.58	3.95	3.95	1.58	314.11
17	8.37	2.68	2.68	4.21	4.21	1.68	313.03
18	8.58	2.79	2.79	4.47	4.47	1.79	313.03
19	8.79	2.89	2.89	4.74	4.74	1.89	314.09
20	9.00	3.00	3.00	5.00	5.00	2.00	314.09

## 4 结论

1) 以铝合金的元素组成为输入向量, 抗拉强度为目标变量, 采用 RF、ET、Bagging、Adaboost 建立了不同的机器学习算法模型, 采用皮尔逊相关系数 (R) 和平均绝对误差对不同机器学习算法模型的性能进行了评价, 结果表明: RF 模型具有最佳的预测性能,  $R = 0.89$ ,  $MAE = 40.33$ 。

2) 采用基于博弈论的 SHAP 方法对 RF 模型进行的解释, 结果表明: Ti 元素含量对铝合金抗拉强度的预测起正向作用, Ti 元素含量越高, 抗拉强度值越大; Mg 元素、Cu 元素含量对铝合金抗拉强度的预测作用并不明显; Zn 元素、Ce 元素、Y 元素含量对铝合金抗拉强度的预测起负向作用, 即元素含量越大, 抗拉强度值越小, 特征重要性从大到小分别为  $Ti > Mg > Cu > Zn > Ce > Y$ 。

### [参考文献]

- [1] 闵承鑫. 8.5% Zn 含量 7xxx 铝合金薄板的制备工艺及力学性能研究[D]. 鞍山: 辽宁科技大学, 2017.
- [2] 马思怡, 张伟健, 苏睿明, 等. 7xxx 系铝合金回归再时效的研究现状[J]. 有色金属科学与工程, 2022(13): 38 - 50.
- [3] WILLIAMS J C, STARKE JR E A. Progress in structural materials for aerospace systems [J]. Acta materialia, 2003, 51(19): 5775 - 99.
- [4] TANG J, ZHANG H, TENG J, et al. Effect of Zn content on the static softening behavior and kinetics of Al-Zn-Mg-Cu alloys during double-stage hot deformation [J]. Journal of Alloys and Compounds, 2019, 806: 1081 - 96.
- [5] 林洪, 刘利. 双级时效对 7XXX 系铝合金力学性能及晶间腐蚀性能的影响[J]. 锻压技术, 2015(40): 122 - 125.
- [6] VON LILIENFELD O A. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties [J]. International Journal of Quantum

Chemistry, 2013, 113(12): 1676 - 1689.

- [7] 赵婉辰, 郑晨, 肖斌, 等. 基于 Bayesian 采样主动机器学习模型的 6061 铝合金成分精细优化[J]. 金属学报, 2021(57): 797 - 810.
- [8] 杜杨. 机器学习预测铝合金搅拌摩擦焊孔洞缺陷形成及其补焊的仿真[D]. 天津: 天津大学, 2020.
- [9] 曹鑫宇. 基于机器学习方法的 7XXX 系铝合金设计及抗应力腐蚀性能与机理研究[D]. 成都: 西南交通大学, 2020.
- [10] 贵星卉, 钟鼓, 李虎田, 等. 合金元素对 7xxx 系铝合金力学性能和晶间腐蚀性能的影响[J]. 有色金属工程, 2021(11): 22 - 28.
- [11] LI J, ZHANG Y, CAO X, et al. Accelerated discovery of high-strength aluminum alloys by machine learning [J]. Communications Materials, 2020, 1(1): 73.
- [12] BREIMAN L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123 - 140.
- [13] BREIMAN L. Random Forests [J]. Machine Learning, 2001, 45(1): 5 - 32.
- [14] GEURTS P, ERNST D, WEHENKEL L. Extremely randomized trees [J]. Machine Learning, 2006, 63(1): 3 - 42.
- [15] FREUND Y, SCHAPIRE R E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [J]. Journal of Computer and System Sciences, 1997, 55(1): 119 - 139.
- [16] ZHAO Z, ANAND R, WANG M. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform; proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), F 5 - 8 Oct. 2019 [C].
- [17] LUNDBERG S M, LEE S-I. A unified approach to interpreting model predictions [J]. Advances in neural information processing systems, 2017, 30.

(下转第 82 页)

## The Calculation and Selection of Industrial Circulating Water Pumps and Pipelines

HOU Chen-ru, NAN Jun-fang, ZHANG Ye-xin, XUE Jian-sen, SU Jiang-feng

**Abstract:** During the construction of a certain factory building, it is necessary to equip a cooling system for high-temperature equipment. In order to transport the circulating water with a flow of  $37 \text{ m}^3/\text{h}$  to the cooling bed and complete the cooling task of the material tray, it is necessary to select a suitable pump and pipeline. The pipeline route is preset according to the layout in the workshop. Through relevant theoretical calculation, the inlet pipe specification of the pump is selected  $\Phi 102 \times 4 \text{ mm}$ ; the outlet pipe size is  $\Phi 73 \times 4 \text{ mm}$ ; the pump model is ISGB65-12.5 (I). Finally pipeline design and pump selection are completed.

**Key words:** circulating water; pipeline design; theoretical calculation; pump selection



---

(上接第 71 页)

## Prediction of Aluminum Alloy Properties Based on Different Machine Learning Algorithms

LI Ting

**Abstract:** Aluminum alloys are widely used in aerospace and transportation fields due to their high strength and low weight. In this paper, with aluminum alloy components as the input vector and tensile strength as the target variable, four different machine learning algorithm models, RF, ET, Bagging and Adaboost, are established. The results show that RF model has the best prediction performance,  $R = 0.89$ ,  $MAE = 40.33$ . The content of Ti plays a positive role in predicting the tensile strength of aluminum alloys. The higher the content of Ti, the greater the tensile strength. The prediction of tensile strength of aluminum alloy with Mg and Cu content is not obvious. The content of Zn element, Ce element and Y element plays a negative role in predicting the tensile strength of aluminum alloy. The element content is great, that the tensile strength value is small. The importance of the characteristics is  $Ti > Mg > Cu > Zn > Ce > Y$ .

**Key words:** machine learning; aluminum alloy; tensile strength; performance prediction

