

## 机器学习辅助的 WC-Co 硬质合金硬度预测\*

宋睿 刘雪梅<sup>†</sup> 王海滨 吕皓 宋晓艳<sup>‡</sup>

(北京工业大学材料科学与工程学院, 新型功能材料教育部重点实验室, 北京 100124)

(2024 年 2 月 22 日收到; 2024 年 4 月 15 日收到修改稿)

硬度是硬质合金材料的一项典型代表性能, 受多种因素的影响且各因素间常存在关联关系. 本文旨在获得 WC-Co 硬质合金硬度的关键影响因素并实现硬度的高通量预测. 建立了以硬质合金硬度为目标变量, 以原料成分、烧结工艺和烧结体表征信息为特征的数据集; 通过对特征的皮尔逊相关系数和 SHAP 分析, 发现 WC 晶粒尺寸和 Co 含量对硬质合金硬度的影响最为显著. 基于机器学习的支持向量机、多项式回归、梯度提升决策树、随机森林等算法, 分别构建了硬质合金硬度预测模型. 采用 10 折交叉验证方法对模型进行定量评估, 结果表明梯度提升决策树算法模型具有最高的精度和较强的泛化能力, 是最适合硬质合金硬度预测的机器学习方法. 基于优选模型的高通量预测数据, 采用多项式回归算法确定了硬质合金硬度与 Co 含量和 WC 晶粒尺寸之间的定量关系, 预测准确率达到 0.946. 本研究为硬质合金性能的准确高效预测提供了数据驱动方法, 可为高性能硬质合金材料的设计研发提供重要参考.

**关键词:** 硬质合金, 硬度, 机器学习, 高通量预测**PACS:** 62.20.Qp, 62.25.-g, 74.70.Ad**DOI:** 10.7498/aps.73.20240284

## 1 引言

硬质合金是由难熔金属化合物 (如 WC, TiC, TaC 等) 和过渡族金属 (Co, Ni, Fe 等) 组成, 通过粉末冶金方法制备的复合材料. 硬质合金广泛应用于切削工具、矿山工具、模具等<sup>[1-5]</sup> 的制备, 已成为工业领域不可或缺的重要材料<sup>[6]</sup>. 硬质合金的使用范围和使用寿命在很大程度上取决于其力学性能, 如强度、硬度、韧性及耐磨、耐蚀等性能. 其中, 硬度作为一项典型的标准性能, 在硬质合金材料质量控制中起着重要作用<sup>[7,8]</sup>. 已有研究表明, 硬质合金硬度的影响因素很多, 包括硬质合金成分 (如 Co 含量、碳含量、晶粒长大抑制剂成分及含量等)、烧结工艺 (如烧结温度、保温时间、烧结压力、升温速率等) 及制备得到硬质合金微观组织表征参量 (如晶粒尺寸、WC 晶粒邻接度、Co 平均自由程等)<sup>[9-13]</sup>.

针对上述影响因素, 已通过实验方法获得了一定的结果, 为硬质合金的开发提供了很好的基础.

然而, 仅靠实验研究无法完全满足现代制造业所需的新型硬质合金开发和个性定制的要求. 以纳米晶硬质合金研究为例: 根据 Hall-Petch 公式推断<sup>[14]</sup>, 当 WC 晶粒尺寸降至纳米尺度 ( $<100$  nm) 时硬质合金的硬度会明显增大. 然而, 烧结过程中粉末原料的致密化和 WC 晶粒长大同时进行, 全致密纳米晶硬质合金的制备非常困难. 现有研究表明, 相比于超细晶硬质合金, 近全致密纳米晶硬质合金的硬度并没有大幅的提高<sup>[15,16]</sup>. 因此, 亟需阐明全致密的纳米晶硬质合金是否具有人们期待的高硬度, 为高性能硬质合金的开发提供研发方向. 尤为重要, 影响硬质合金硬度的因素通常存在关联关系, 如不同 Co 含量的硬质合金, 与之匹配的烧结工艺参数不同<sup>[17-19]</sup>; 随着烧结温度的降低和晶粒长大抑制剂的添加, WC 晶粒尺寸减小;

\* 国家自然科学基金 (批准号: 52271085, 92163107, 52171061) 资助的课题.

<sup>†</sup> 通信作者. E-mail: liuxuemei@bjut.edu.cn

<sup>‡</sup> 通信作者. E-mail: xysong@bjut.edu.cn

随着 Co 含量的增大, WC 晶粒邻接度减小、Co 平均自由程增大; 高 Co 含量硬质合金可在较低的烧结温度下致密化, 可降低 W 和 C 原子在 Co 液相中的溶解/析出速度, 从而降低 WC 晶粒尺寸等. 因此, 如何选择关键的影响因素, 获得它们的优化匹配是制备高硬度硬质合金的关键<sup>[20,21]</sup>. 目前结合经验和实验的方法, 凭实验“试错”来探索适当的微观组织结构组合获得合适硬度的硬质合金方法, 可控性差, 费时费力. 因此, 急需寻找一个能够有效快速识别硬质合金硬度的策略, 为硬质合金的研发提供指导.

机器学习是研究计算机如何模拟人类学习行为以自动获取知识, 从而不断改善自身性能的一门学科. 机器学习作为一种新兴的学习方式, 它从已有的数据中学习以产生可靠的、可重复的决策和结果, 在材料力学性能预测方面发挥了重要作用<sup>[22-29]</sup>. 然而, 目前在硬质合金材料中的应用较少, 仅有的报道基于机器学习建立了一种预测 WC-10%Co-0.5%Cr<sub>3</sub>C<sub>2</sub> 硬质合金 (含量单位为质量分数) 抗弯强度的可重复性工艺预测模型, 基于文献收集数据, 实现了通过合金成分、物理特性等对该硬质合金的密度、矫顽力、横向断裂强度等力学性能的预测, 均获得了较好的预测效果<sup>[30,31]</sup>. 由此可见, 机器学习方法对硬质合金力学性能预测是完全可行性的.

本研究的目的是建立适用于硬质合金材料、具有更高预测准确度和适用性的硬度预测模型. 以应用最为广泛的 WC-Co 硬质合金为例, 基于从文献中收集、清洗等得到的实验数据, 通过特征筛选选择影响硬质合金硬度的关键特征. 通过 4 种机器学习算法的模型, 在参数优化的基础上对 WC-Co 硬质合金的硬度进行了预测, 并经过 10 次 10 折交叉验证及多种评价指标对模型在测试集上的预测结

果进行评估, 得到综合预测误差最小的机器学习模型; 最后, 为获得硬质合金硬度与关键影响特征间的定量关系, 以预测精度最高的机器学习模型预测结果为原始数据, 采用多项式回归方法进行硬质合金硬度的预测, 为高硬度硬质合金的研发提供参考.

## 2 方法

图 1 是硬质合金硬度机器学习 (machine learning, ML) 模型的构建流程, 主要包括机器学习数据集的构建、特征筛选、选取特征的标准化处理及最优算法模型的选择、预测数据的可解释性模型建立及 WC-Co 硬质合金硬度的高通量预测.

### 2.1 原始数据集的建立与数据清洗

通过知网、Science Direct、Web of Science 等公开平台正式出版的文献, 获取了 360 条关于硬质合金硬度的数据. 在初始特征选择时, 本研究将影响 WC-Co 硬质合金硬度的输入特征分为化学成分、制备工艺参数、硬质合金组织和性能等表征参数. 硬度 (hardness) 作为目标变量, 其范围在 800—2319 kgf/mm<sup>2</sup>; 化学成分包括 C 含量 (C content)、O 含量 (O content)、Co 含量 (Co content), 相应取值范围 (质量分数) 分别为 5.34%—6%, 0.18%—0.67% 和 3%—15%; 制备工艺参数包括烧结温度 (sintering temperature)、压力 (pressure)、保温时间 (holding time), 对应取值范围分别为 1100—1450 °C, 0—120 MPa 和 5—100 min; 硬质合金组织和性能的表征参数包括 WC 的晶粒尺寸 (WC grain size) 以及矫顽力 (coercive force), 其范围分别位于 0.096—10.8 μm, 2.1—41.2 kA/m. 鉴于部

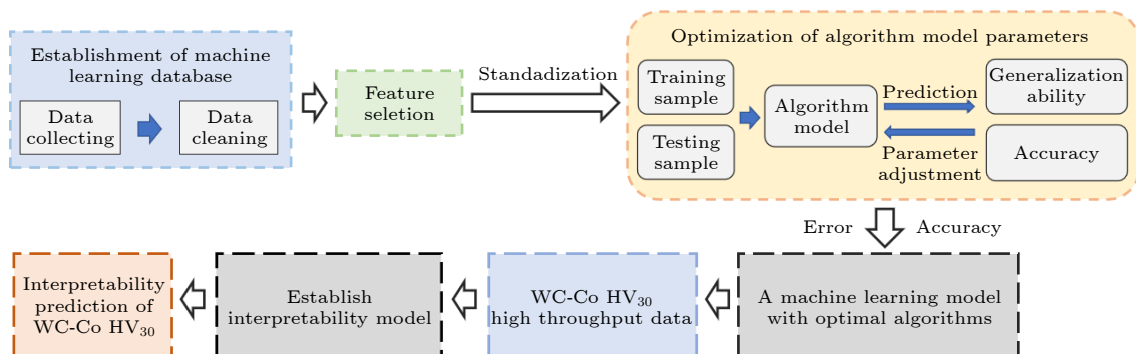


图 1 WC-Co 硬质合金硬度预测的机器学习模型构建流程图

Fig. 1. Hardness prediction workflow of WC-Co cemented carbides based on ML.

分数据特征不全的问题, 对不同实验组数值变化不大的 C 含量、O 含量进行填补, 填补值为各实验中该特征的平均值; 对缺失变化相对较大的特征 (如 Co 含量、WC 晶粒尺寸、保温时间、烧结温度、压力) 的数据组进行删除; 由于矫顽力的数据量较少, 对含有矫顽力表征参数的数据组, 保持不变, 即对缺少矫顽力的数据组保持矫顽力参数空缺。

## 2.2 特征筛选

特征筛选是一个非常重要的数据预处理过程. 为剔除冗余和硬质合金硬度不相关的特征, 提高机器学习模型的精确, 采用各特征之间的皮尔逊相关系数 (pearson correlation coefficient) 和 SHAP (shapley additive explanations)<sup>[32,33]</sup> 分析方法, 筛选出影响硬质合金硬度的关键特征。

皮尔逊相关系数可以分析和识别具有高相关性的特征, 从而消除具有高度相关或线性相关关系的特征, 以保证模型的稳定性; 而对于具有非简单线性关系的特征变量, 通常通过计算特征的重要性来进行筛选. 基于算法模型 (如决策树、随机森林等) 的 SHAP 分析是特征重要性预测的常用方法. 对于特定样本的预测值, SHAP 值就是该样本中每个特征对预测值的贡献<sup>[34]</sup>. SHAP 值服从以下等式:

$$y_i = y_{\text{base}} + f(x_{i,1}) + f(x_{i,2}) + \cdots + f(x_{i,j}), \quad (1)$$

其中,  $y_i$  为模型对第  $i$  个样本的预测值;  $y_{\text{base}}$  为整个模型的基线 (通常是所有样本的目标变量的均值);  $x_i$  为第  $i$  个样本,  $x_{i,j}$  为第  $i$  个样本的第  $j$  个特征;  $f(x_{i,j})$  为  $x_{i,j}$  的 SHAP 值. SHAP 值为正时表示该特征的取值倾向于增大因变量的值, SHAP 值大小即为增加的量; SHAP 值为负时表示该特征的取值倾向于减小因变量的值, SHAP 值大小即为减小的量. SHAP 分析不仅可消除不同算法模型得到结果的不一致性, 还可以表征各自变量对因变量的影响趋势和程度。

## 2.3 机器学习模型算法的选择

机器学习算法的选择是机器学习模型构建的关键. 由于本研究收集的影响 WC-Co 硬度的特征均为连续型变量, 因此选择可预测连续输出量的回归算法. 本文采用支持向量机回归 (support vector regression, SVR)<sup>[35]</sup>、多项式回归 (polynomial regression, PR)<sup>[36]</sup>、梯度提升决策树 (gradient boost-

ing decision tree, GBDT)<sup>[37]</sup>、随机森林 (random forest, RF)<sup>[38]</sup> 等常用的 4 种机器学习算法进行硬质合金硬度的预测. 以均方误差 (MSE)((2) 式)、平均绝对误差 (MAE)((3) 式) 和代表模型预测准确率的决定系数 ( $R^2$ )((4) 式) 来评估机器学习模型的优劣。

$$\text{MSE} = \frac{1}{N} \sum_{m=1}^N (y'_m - y_m)^2, \quad (2)$$

$$\text{MAE} = \frac{1}{N} \sum_{m=1}^N |y_m - y'_m|, \quad (3)$$

$$R^2 = \frac{\sum_{m=0}^{N-1} (y_m - \bar{y}_m)^2}{\sum_{m=0}^{N-1} (y_m - y'_m)^2}, \quad (4)$$

其中,  $y_m$  为第  $m$  个样本的真实值,  $y'_m$  为第  $m$  个样本的预测值,  $\bar{y}_m$  为所有真实值的平均值. 机器学习模型预测结果的 MSE 和 MAE 越小、 $R^2$  越大, 表明该模型越适合硬质合金硬度的预测。

## 2.4 机器学习算法模型参数的优化

SVR 算法模型的主要参数包括惩罚系数、核函数以及核函数系数等, 而 PR 算法模型的主要参数为多项式回归函数的阶数. 针对主要参数数量较少的 SVR 和 PR 算法模型, 采用网格寻找基础参数, 再经 10 次 10 折交叉验证的方法确定其最优参数。

对于主要参数较多的 GBDT 与 RF 算法模型, 鉴于网格搜索确定参数需要的时间较长, 采用手动调参得到优化参数. 即针对某一参数, 在给定的范围内, 结合模型测试集的泛化误差随参数的变化趋势, 优化获得基础参数. 对具有多个基础参数组合的算法模型进行 10 次 10 折交叉验证, 如果该算法模型的准确率满足要求, 则选择该参数组合作为算法模型的优化参数; 相反, 如果该算法模型的准确率不满足要求, 则重新调整参数范围, 进行参数的调整, 直至参数组合满足要求. 调参过程中模型测试集的泛化误差是用来衡量一个学习机器推广未知数据的能力, 即根据从样本数据中学习到的规则能够应用到新数据的能力, 是评估模型对于未知数据的预测能力的重要指标. 泛化误差由偏差 (bias)、方差 (variance, 简称 Var) 及噪声组成. 噪声表示当前任务上学习算法所能达到的期望泛化误差的下界, 主要来自于“数据采集”的过程 (如试验设备



产生的随机偏差、实验中其他不可控因素的干扰等), 因此, 噪声无法通过模型参数来消除, 通常在数据清洗阶段去除异常数据来减小噪声. 偏差表示模型中预测值和真实值之间的差异, 方差用来衡量算法模型的稳定性. 算法模型预测越准确、越稳定、泛化能力越强, 其偏差和方差越低.

### 3 结果与讨论

#### 3.1 影响硬质合金硬度关键特征的选取

图 2 所示是影响硬质合金硬度特征之间的皮尔逊相关系数图. 图中颜色越深表明两特征之间存在极强的负向相关性 (最小相关性系数为-1), 颜色越浅表明两特征存在极强的正向相关性 (最大相关性系数为 1).

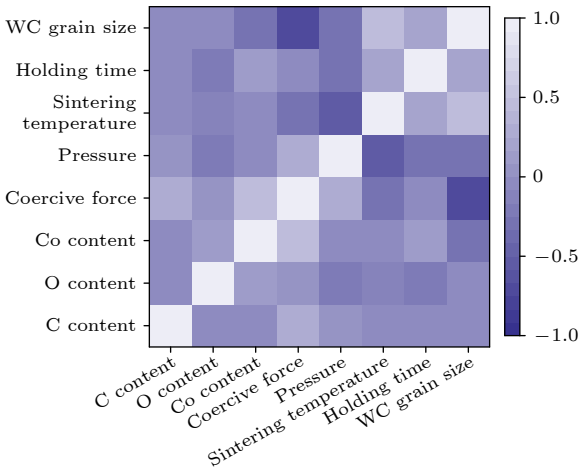


图 2 影响硬质合金硬度特征之间的皮尔逊线性相关系数  
Fig. 2. Pearson linear correlation coefficient of among the influence features on the hardness of cemented carbides.

可以看出, 矫顽力与 WC 晶粒尺寸的皮尔逊相关系数为-0.73, 表明两者之间存在着高度负向线性相关性. 综合考两个特征的使用频率, 删去矫顽力这一特征, 保留 WC 晶粒尺寸特征.

进一步, 基于随机森林模型, 得到各特征对硬质合金 WC 晶粒尺寸的 SHAP 值及其绝对值的平均值如图 3 所示. 可以看出, WC 晶粒尺寸对硬度的影响最大, 其次是 Co 含量, 之后是压力、烧结温度等. 另外, 图 3(a) 中每个采样点的颜色代表该特征的数值大小, 红色代表特征的数值高, 而蓝色意味着数值低. 可得到特征对硬度的影响规律. 以 WC 晶粒尺寸为例, 大 WC 晶粒尺寸 (红色点所示) 可降低硬质合金硬度, 而小 WC 晶粒尺寸 (蓝色点所示) 可提高其硬度, 且 WC 晶粒尺寸越小, 对硬度的影响程度越大.

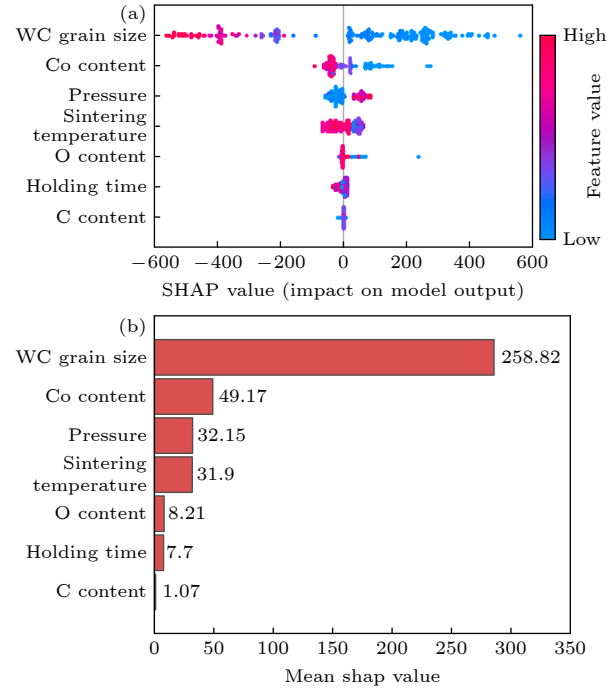


图 3 目标变量为硬度时各特征 SHAP 值 (a) 和各特征平均 SHAP 的绝对值 (b) 的排序

Fig. 3. SHAP values (a) and the absolute value of average SHAP (b) of each feature with target variable of hardness.

综上所述, WC 晶粒尺寸和 Co 含量是影响硬质合金硬度的关键特征. 因此, 选择 WC 晶粒尺寸和 Co 含量作为关键特征, 进行硬质合金硬度的高通量预测.

#### 3.2 算法模型参数的优化

由于 GBDT 模型与 RF 模型的参数大部分相同, 如弱学习器数量 (number of estimator)、树的深度 (max depth)、叶子节点最少样本数 (min sample leaf) 以及内部节点再划分所需最小样本数 (min sample split) 等. 但有少数参数不同, 如 GBDT 需要查找适合的子采样比例 (subsample)、学习率 (learning rate) 以及损失函数, 而 RF 需要调节合适的最大特征数 (max feature). 考虑到二者使用的参数优化方式相同, 因此本文以 GBDT 模型的部分重要参数优化为例, 介绍手动调参的过程. 具体结果如图 4 所示.

图 4(a) 展示了弱学习器数量的调参范围为 0—200 时, GBDT 算法模型测试集的决定系数、偏差和方差. 可以看出, 当弱学习器数量为 70 时, 模

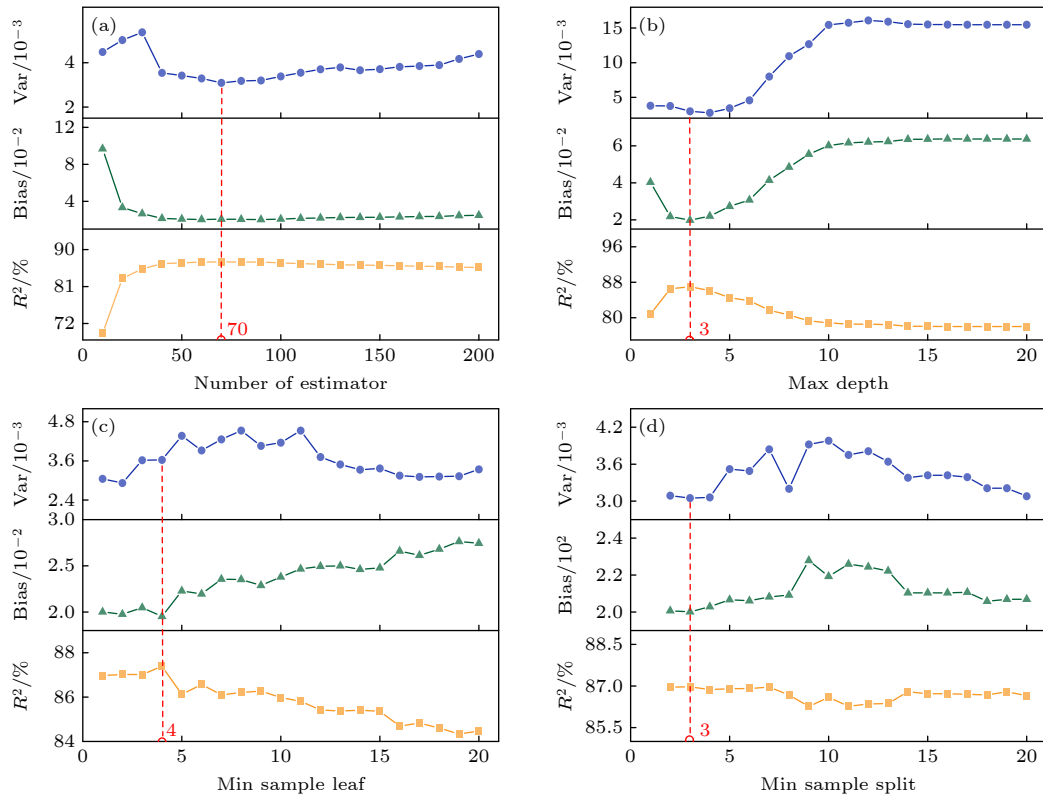


图 4 典型参数对 GBDT 算法模型的测试集准确率 ( $R^2$ )、偏差 (Bias) 和方差 (Var) 的影响 (a) 弱学习器数量; (b) 树的最大深度; (c) 叶子节点最少样本数; (d) 内部节点再划分所需最小样本数

Fig. 4. Performance of typical parameters on the testing set in terms of accuracy ( $R^2$ ), bias (Bias) and variance (Var) based on GBDT model: (a) Number of estimator; (b) max depth; (c) min sample leaf; (d) min sample split.

型具有较高的  $R^2$  值, 达到了 87.2%, 同时具有较低的偏差值 0.0207 和最低的方差值 0.00309. 也就是说, 此时模型具有良好预测准确性、稳定性和泛化能力. 根据图 4(b)—(d) 可以看出, 树的深度为 3、叶子节点最少样本数为 4、内部节点再划分所需最小样本数为 3 时, GBDT 模型具有最高的  $R^2$  值、较低的偏差和方差值. 将上述优化参数组合作为 GBDT 模型的基础参数. 基于上述优化参数组合的 GBDT 算法模型, 对数据进行 10 次 10 折交叉验证, 结果表明模型的准确率超过 87%, 由此确定上述基础参数组合可作为 GBDT 算法模型的最优参数, 进行硬质合金硬度的预测. RF 算法模型的最优参数采用相同的方法得到.

### 3.3 模型效果及评价

机器学习模型构建过程中, 训练集和测试集划分按照一定数据比例随机划分、进而采用 10 折交叉验证的方法, 确定数据划分的最优比例. 考虑到模型训练需要足够多的数据, 因此训练集和测试集的数据比例分别选择了 9:1, 8:2, 7:3 和 6:4 进行

模型的训练. 结果表明, 训练集和测试集数据比例在 7:3 时, 4 种机器学习算法模型的测试集  $R^2$  相对最高. 因此, 将预处理后的数据集按照 7:3 的比例随机分成训练集和测试集. 采用参数优化后的 4 种机器学习算法模型对训练集进行学习, 并对测试集进行预测, 结果如图 5 和图 6 所示. 图中的绿色和红色的实心圆分别代表训练集和测试集的数据, 虚线上数据表明硬质合金硬度的真实值和预测值完全吻合, 数据越靠近该虚线, 表明预测值和真实值之间的绝对误差越小.

由图 5 可知, 4 种算法模型得到训练集的数据大多数落在硬质合金硬度的真实值和预测值完全吻合的虚线周围, 预测效果较好. 相比较而言, RF 算法的  $R^2$  值最大达到 0.95 (图 5(d)), 训练效果最好; 其次是 GBDT 算法和 SVR 算法,  $R^2$  值分别为 0.93 和 0.87 (图 5(a), (c)), 而 PR 算法训练效果相对较差,  $R^2$  值仅为 0.86 (图 5(b)). 对于测试集的预测效果 (图 6) 而言, GBDT 算法的预测精度最高, 其次是 RF 算法和 SVR 算法, PR 模型的预测效果最差.

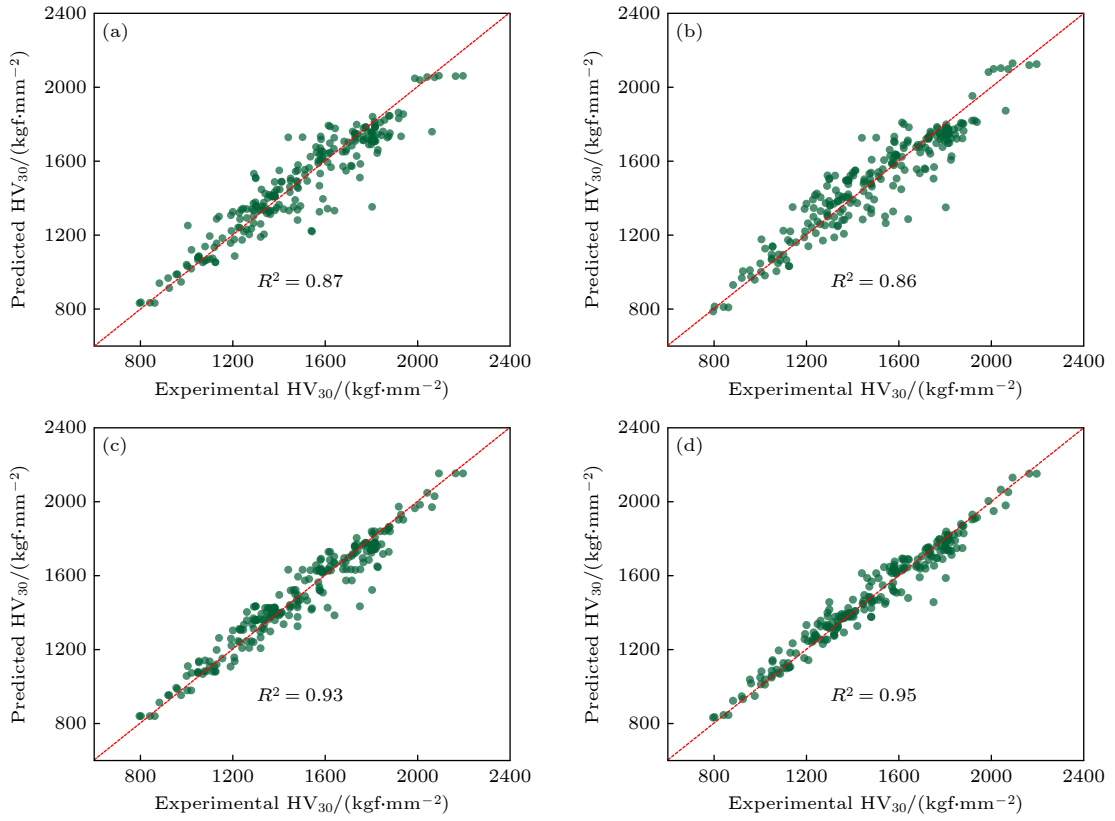


图 5 四种算法模型训练集学习效果 (a) SVR 算法; (b) PR 算法; (c) GBDT 算法; (d) RF 算法

Fig. 5. Performance of four models on training set: (a) SVR algorithm; (b) PR algorithm; (c) GBDT algorithm; (d) RF algorithm.

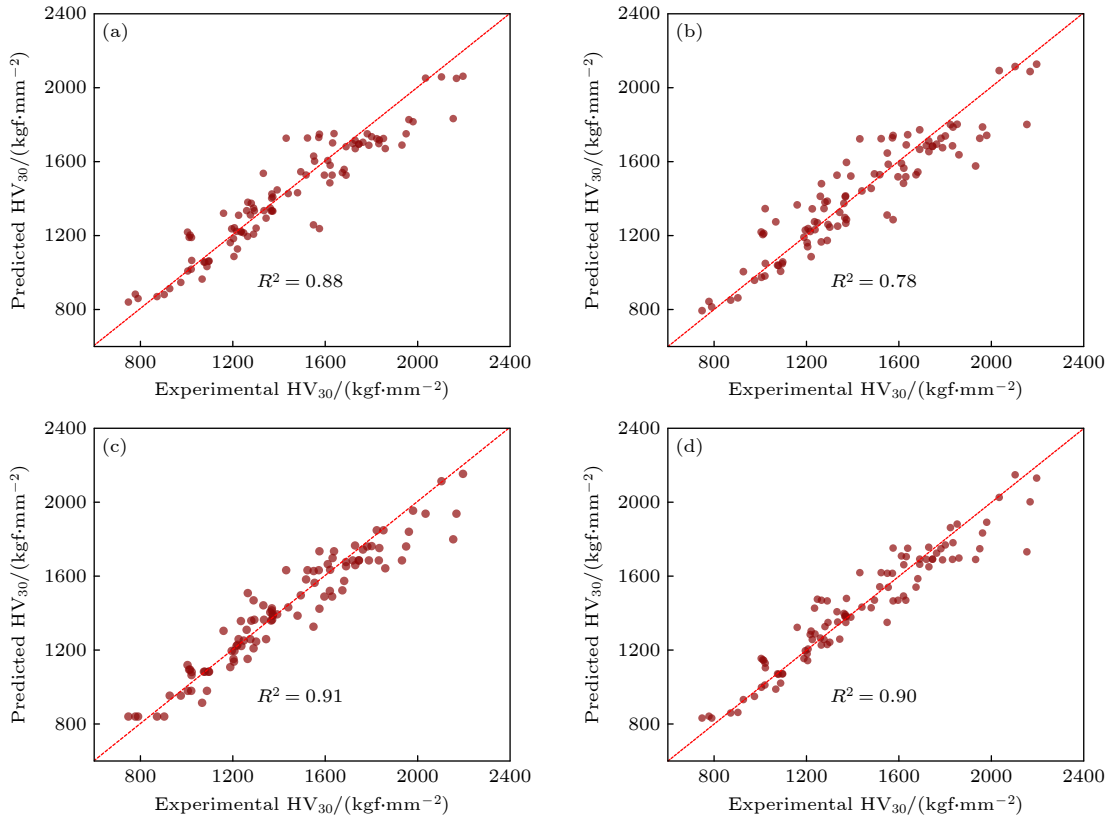


图 6 四种算法模型测试集学习效果 (a) SVR 算法; (b) PR 算法; (c) GBDT 算法; (d) RF 算法

Fig. 6. Performance of four models on testing set: (a) SVR algorithm; (b) PR algorithm; (c) GBDT algorithm; (d) RF algorithm.

为更为全面地评估机器学习算法模型的优劣,选择合适的硬质合金硬度预测模型,对不同算法模型测试集的 MSE 和 MAE 以及经 10 次 10 折交叉验证后得到测试集的  $R^2$  进行了分析,结果如图 7 所示.可以看出,4 种算法模型的  $R^2$  值均超过 77%.其中,GBDT 算法模型的  $R^2$  值最高,为 87.3%;MSE 与 MAE 最低,分别为 0.11 和 0.25.因此,本研究优先选择误差最低、准确率最高的 GBDT 模型作为预测 WC-Co 硬质合金硬度的最佳模型.

### 3.4 模型预测结果

#### 3.4.1 以 GBDT 算法构成的机器学习模型的预测效果

图 8 所示是 WC-Co 硬质合金硬度随 WC 晶粒尺寸及 Co 含量的变化.可以看出,随着 WC 晶粒尺寸及 Co 含量的增大,硬度整体呈现出下降的趋势.然而实验数据有限,尤其是在一些难制备材

料或目前关注不多区域的数据存在空缺(图 8(a)),例如,在 WC 晶粒尺寸小于 100 nm 区域、低 Co (低于 6%) 区域以及 WC 晶粒尺寸大于 5  $\mu\text{m}$  且 Co 含量高于 10% 的区域等.采用具有优化参数 GBDT 算法的机器学习模型进行高通量预测得到 1313 种 WC-Co 硬质合金的硬度,获得硬度随 WC 晶粒尺寸及 Co 含量的变化如图 8(b) 所示.可以看出,预测得到硬质合金的硬度随着 Co 含量和 WC 晶粒尺寸的变化趋势与原始数据基本一致,体现了 GBDT 算法模型预测的准确性.同时预测结果拓展了实验难以测定区域的数据.

然而,GBDT 模型是通过迭代拟合残差来逐步改进预测结果.每一轮迭代,都会有一个新的弱学习器被训练来拟合上一轮模型的残差,然后将上一轮模型的预测结果与当前的弱学习器的预测结果加权组合,得到当前轮的预测结果.因此,GBDT 的学习效果受原始数据影响很大.由图 8(a) 所示的原始数据可知,硬质合金硬度存在较高或

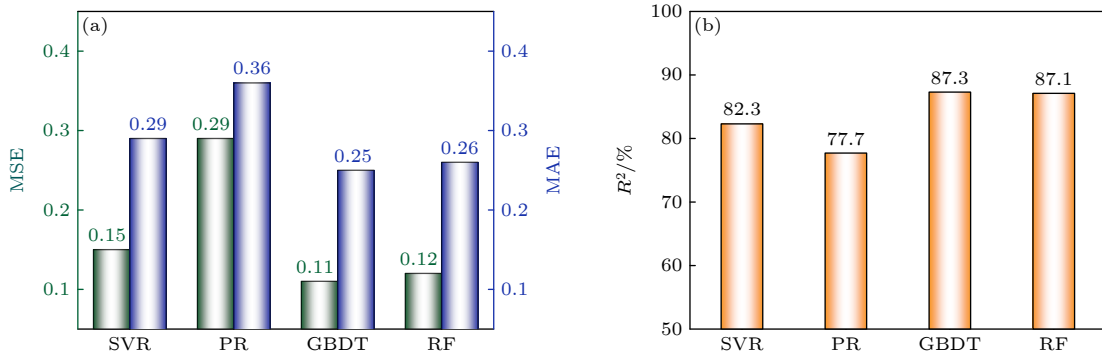


图 7 不同机器学习算法模型测试集效果对比 (a) MSE 和 MAE; (b) 经 10 次 10 折交叉验证得到的  $R^2$

Fig. 7. Performance of different machine learning algorithms on testing set: (a) MSE and MAE; (b) and  $R^2$  score by 10-fold cross-validation.

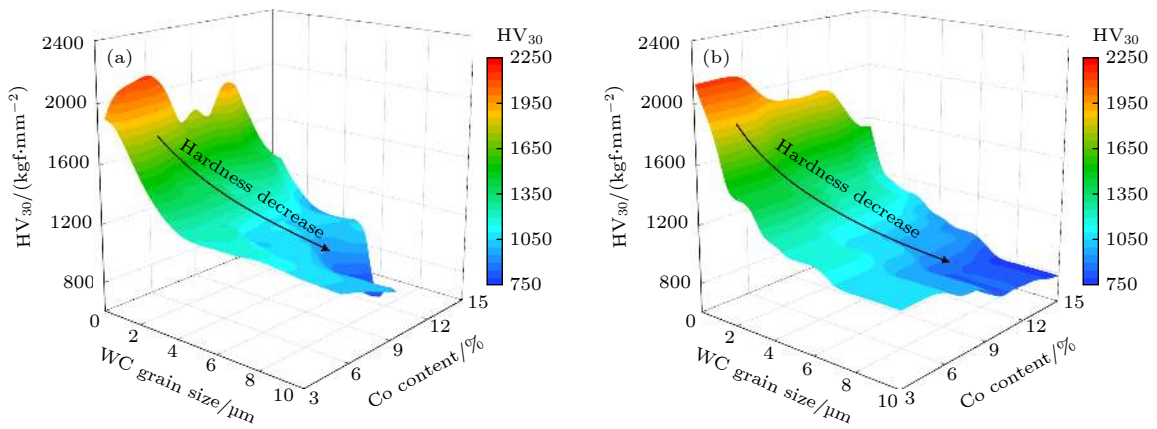


图 8 硬质合金硬度随 WC 晶粒尺寸和 Co 含量的变化 (a) 原始数据; (b) GBDT 模型预测

Fig. 8. Hardness of cemented carbides as a function of WC grain size and Co content: (a) Original data; (b) data predicted by GBDT model.

较低值,使数据分布的连续性较差,导致GBDT算法模型的预测结果存在连续性差的区域;此外GBDT模型属于黑箱模型,无法给出一个准确的表达式来描述硬度和WC晶粒尺寸、Co含量之间的关系规律,不利于后期针对高硬度硬质合金开发进行指导。

### 3.4.2 以GBDT模型预测结果作为输入的机器学习模型预测效果

为实现硬质合金硬度的可解释性预测,以GBDT算法的机器学习模型得到的高通量预测结果为原始数据,通过PR算法模型获得的目标变量硬度与WC晶粒尺寸和Co含量之间的量化关系为

$$\begin{aligned} HV_{30} = & 2156.12 - 388.70d + 6.41\omega + 45.29d^2 \\ & + 7.62d\omega - 6.41\omega^2 - 1.62d^3 - 0.90d^2\omega \\ & + 0.18d\omega^2 + 0.20\omega^3, \end{aligned} \quad (5)$$

其中 $HV_{30}$ 为硬质合金硬度,单位为 $\text{kgf}/\text{mm}^2$  ( $1 \text{ kgf}/\text{mm}^2 = 9.80665 \text{ MPa}$ ); $d$ 为WC晶粒尺寸,单位为 $\mu\text{m}$ ;  $\omega$ 表示Co含量(质量分数),单位为%。

图9(a)所示为PR算法模型训练集和测试集训练与预测效果的评估。可以看出,训练集和测试集的MSE均小于0.1,表明了模型良好的学习能力和预测能力;训练集和测试集的MAE及MSE的差距较小,说明该模型预测结果不存在过拟合。另外,由图9(b)可知,PR算法模型测试集预测准确率达到0.946,表明PR算法模型具有很好的预测结果。

图10是采用PR算法模型对WC-Co硬质合金的硬度进行高通量预测的结果。预测范围为: $0.05 \mu\text{m} < d < 10 \mu\text{m}$ ,  $3\% < \omega < 15\%$ 。由硬度、WC晶粒尺寸、Co含量构成的三维立体图,如图10(a)所示。可以看出,硬质合金的硬度 $HV_{30}$ 分布在750—

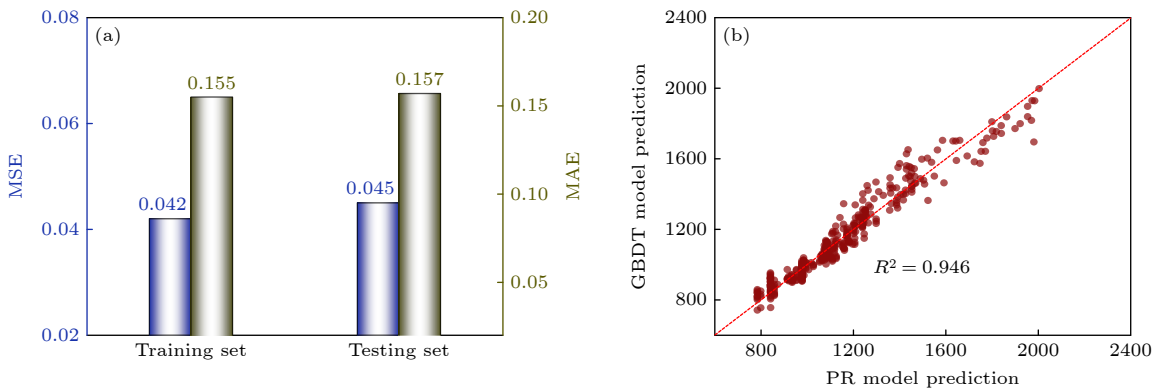


图9 PR算法模型训练及预测效果的评估 (a) 训练集与测试集预测的MAE, MSE; (b) PR算法模型测试集预测准确率  
Fig. 9. Evaluation of the PR model: (a) MSE and MAE for the training and testing sets; (b)  $R^2$  for the testing set.

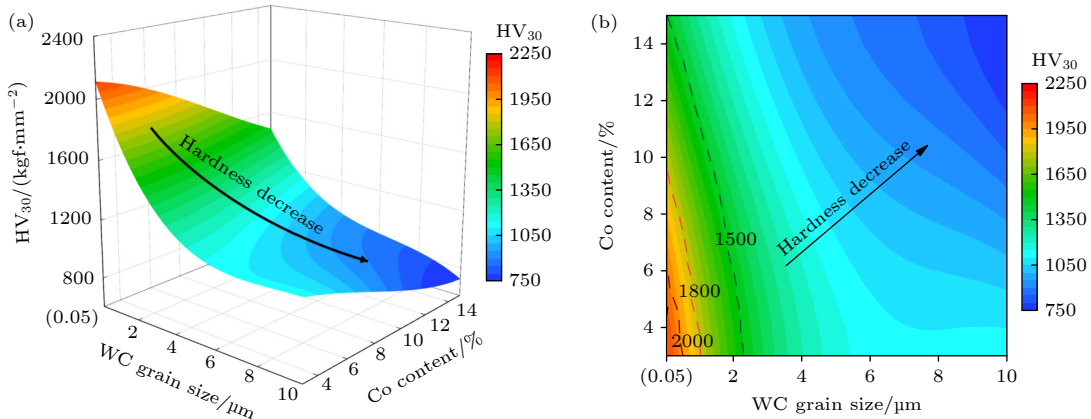


图10 PR算法模型的硬质合金硬度预测结果 (a) 硬度随WC晶粒尺寸、Co含量变化的三维图; (b) 硬度在WC晶粒尺寸和Co含量构成平面上的投影图

Fig. 10. Hardness of cemented carbides predicted by the PR model: (a) Hardness varying with WC grain size and Co content; (b) hardness projection on the plane of WC grain size and Co content.



2125 kgf/mm<sup>2</sup> 之间, 随 WC 晶粒尺寸和 Co 含量的增大呈下降趋势, 且在测定的范围内, 硬度随 WC 晶粒尺寸的变化大于随 Co 含量的变化. 从图 10(b) 可以看出: 在高钴或 WC 晶粒尺寸较大区域 (绿色虚线右侧区域), HV<sub>30</sub> 小于 1500 kgf/mm<sup>2</sup>; 在低钴、小 WC 晶粒尺寸区域 (粉色虚线左侧区域), HV<sub>30</sub> 大于 1800 kgf/mm<sup>2</sup>; 尤其是当硬质合金的 Co 含量和 WC 晶粒尺寸分别低至 6% 和 0.1 μm 时 (红色虚线左侧区域), HV<sub>30</sub> 高达 2000 kgf/mm<sup>2</sup> 以上.

为获得高硬度硬质合金随特征的变化, 对硬度大于 1800 kgf/mm<sup>2</sup> 区域进行了进一步的分析. 硬度随 WC 晶粒尺寸和 Co 含量的变化如图 11 所示. 可以看出, 硬质合金的硬度随着 WC 晶粒尺寸和 Co 含量的增大均呈现下降趋势, 且硬度在 WC 晶粒尺寸为 0.05—1 μm 和 Co 含量在 3%—10% 范围内的变化量相差不大. 硬度的变化速率 (以图 11(b) 所示曲线的斜率  $k$  表示) 如图 12 所示. 随着 WC 晶粒尺寸的增大, 硬度变化速率呈下降趋势, 但随

Co 含量的增大则明显增大, 这表明硬度随 WC 晶粒尺寸和 Co 含量的下降呈非单调变化. WC 晶粒尺寸越小、Co 含量越高, WC 晶粒尺寸以及 Co 含量的变化对硬度的影响越大.

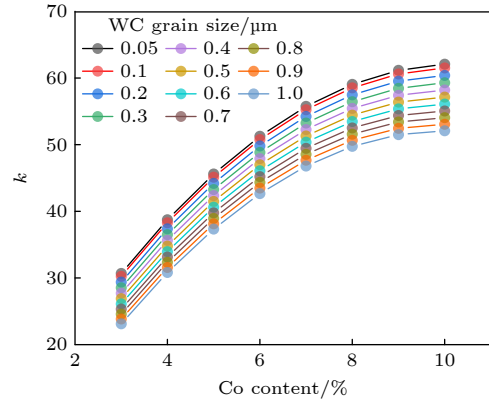


图 12 不同 WC 晶粒尺寸下硬度变化率随 Co 含量的变化  
Fig. 12. Hardness slope with different variables with Co content under different WC grain size.

综上所述, 减小 WC 晶粒尺寸和降低 Co 含量是获得高硬度的关键. 研究结果对设计和制备具有高硬度的 WC-Co 硬质合金具有重要的指导作用.

## 4 结 论

本文通过机器学习用数据库建立、特征抽取、算法模型参数优化和硬质合金硬度的高通量预测, 获得了 WC-Co 硬质合金硬度的关键影响因素及其影响规律, 可为具有高硬度的硬质合金材料设计研发提供指导.

1) 构建了包含 8 种 WC-Co 硬质合金硬度影响因素特征的原始数据集, 筛选出 WC 晶粒尺寸和 Co 含量是影响硬质合金硬度的关键特征;

2) 相比于 SVR, PR 以及 RF 算法模型, GBDT 算法模型对硬质合金硬度的预测准确率最高、MSE 和 MAE 最低, 是最适合预测分析硬质合金硬度的机器学习算法模型;

3) 基于 GBDT 算法模型预测得到的高通量数据, 采用 PR 算法得到了硬质合金硬度与 WC 晶粒尺寸、Co 含量之间的可解释性计算模型, 实现了对 WC-Co 硬质合金硬度的高精度预测;

4) 减小晶粒尺寸和降低 Co 含量是获得硬质合金高硬度的关键, 且相比于 Co 含量, WC 晶粒尺寸的调控对改变硬质合金的硬度更为有效.

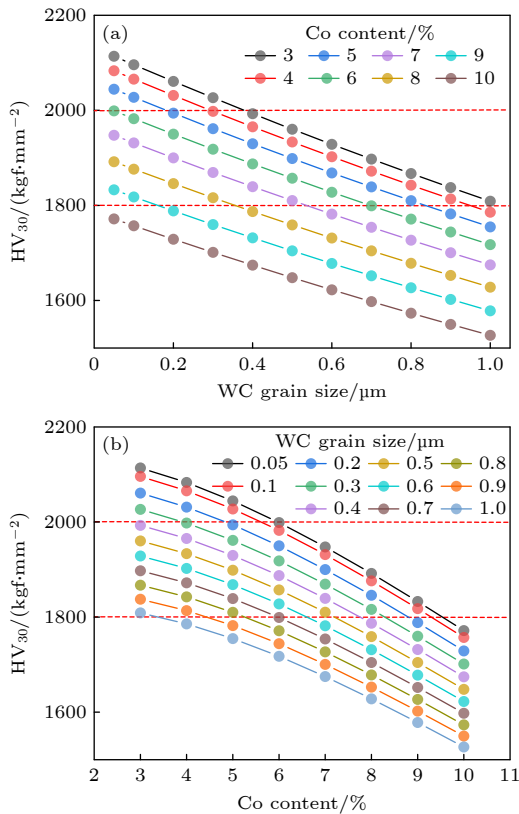


图 11 硬度大于 1800 kgf/mm<sup>2</sup> 区域的硬质合金硬度预测结果 (a) 硬度随 WC 晶粒尺寸的变化; (b) 硬度随 Co 含量的变化

Fig. 11. Prediction of hardness in a range of hardness higher than 1800 kgf/mm<sup>2</sup>: (a) Hardness varying with WC grain size; (b) hardness varying with Co content.

## 参考文献

- [1] Ding Y Z, Ye Y, Li D S, Xu F, Lang W C, Liu J H, Wen X 2023 *Acta Phys. Sin.* **72** 068703 (in Chinese) [丁业章, 叶寅, 李多生, 徐锋, 朗文昌, 刘俊红, 温鑫 2023 物理学报 **72** 068703]
- [2] Useldinger R, Schleinkofer U 2017 *Int. J. Refract. Met. Hard Mater.* **62** 170
- [3] Springs G E 1995 *Int. J. Refract. Met. Hard Mater.* **13** 241
- [4] Ghasali E, Orooji Y, Tahamtan H, Asadian K, Alizadeh M, Ebadzadeh T 2020 *Ceram. Int.* **46** 29199
- [5] Ezquerro B L, Lozada L, Berg H V D, Wolf M, Sánchez J M 2018 *Int. J. Refract. Met. Hard Mater.* **72** 89
- [6] Sun L, Yang T E, Jia C C, Xiong J 2011 *Int. J. Refract. Met. Hard Mater.* **29** 147
- [7] Ding Q J, Zheng Y, Ke Z, Zhang G T, Wu H, Xu X Y, Lu X P, Zhu X G 2020 *Int. J. Refract. Met. Hard Mater.* **87** 105166
- [8] Hu H X, Liu X M, Chen J H, Lu H, Liu C, Wang H B, Luan J H, Jiao Z B, Liu Y, Song X Y 2022 *J. Mater. Sci. Technol.* **104** 8
- [9] Yu S B, Min F L, Ying G B, Noudem J G, Liu S J, Zhang J F 2021 *Mater. Charact.* **180** 111386
- [10] Tang Y Y, Wang S N, Xu F Y, Hong Y K, Luo X, He S M, Chen L Y, Zhong Z Q, Chen H, Xu G Z, Yang Q M 2021 *J. Alloy Compd.* **882** 160638
- [11] Jafari M, Enayati M H, Salehi M, Nahvi S M, Park C G 2014 *Ceram. Int.* **40** 11031
- [12] Wang H, Zeng M Q, Liu J W, Lu Z C, Shi Z H, Ouyang L Z, Zhu M 2015 *Int. J. Refract. Met. Hard Mater.* **48** 97
- [13] Singla G, Singh K, Pandey O P 2014 *Ceram. Int.* **40** 5157
- [14] Liu W H, Wu Y, He J Y, Nieh T G, Lu Z P 2013 *Scripta Mater.* **68** 526
- [15] Liu X M, Song X Y, Wei C B, Gao Y, Wang H B 2012 *Scripta Mater.* **66** 825
- [16] Song X Y, Gao Y, Liu X M, Wei C B, Wang H B, Xu W W 2013 *Acta Mater.* **61** 2154
- [17] Bonache V, Salvador M D, Fernández A, Borrell A 2011 *Int. J. Refract. Met. Hard Mater.* **29** 202
- [18] Fang Z, Maheshwari P, Wang X, Sohn H Y, Griffio A, Riley R 2005 *Int. J. Refract. Met. Hard Mater.* **23** 249
- [19] Fang Z Z, Wang X, Ryu T, Hwang K S, Sohn H Y 2009 *Int. J. Refract. Met. Hard Mater.* **27** 288
- [20] Liu K, Wang Z H, Yin Z B, Cao L Y, Yuan J T 2018 *Ceram. Int.* **44** 18711
- [21] Zhao S X, Song X Y, Liu X M, Wei C B, Wang H B, Gao Y 2011 *Acta Metall. Sin.* **47** 1188 (in Chinese) [赵世贤, 宋晓艳, 刘雪梅, 魏崇斌, 王海滨, 高杨 2011 金属学报 **47** 1188]
- [22] Roy A, Babuska T, Krick B, Balasubramanian G 2020 *Scripta Mater.* **185** 152
- [23] Chanda B, Jana P P, Das J 2021 *Comp. Mater. Sci.* **197** 110619
- [24] George K, Haoyan D, Chanhho L, Samaei A T, Tu P, Maarten J, Ke A, Dong M, Peter K L, Wei C 2019 *Acta Mater.* **181** 124
- [25] Bakr M, Syarif J, Hashem I A T 2022 *Mater. Today. Commun.* **31** 103407
- [26] Ozerdem M S, Kolukisa S 2009 *Mater. Design* **30** 764
- [27] Sun Y, Zeng W D, Han Y F, Ma X, Zhao Y Q, Guo P, Wang G, Dargusch M S 2012 *Comp. Mater. Sci.* **60** 239
- [28] Zhang X Y, Dong R F, Guo Q W, Hou H, Zhao Y H 2023 *J. Mater. Res. Technol.* **26** 4813
- [29] Catal A A, Bedir E, Yilmaz R, Swider M A, Lee C, El-Atwani O, Maier H J, Ozdemir H C, Canadinc D 2024 *Comp. Mater. Sci.* **231** 112612
- [30] Guan Z H, Tian H X, Li N, Long J Z, Zhang W B, Du Y 2023 *Ceram. Int.* **49** 613
- [31] Guan Z H, Li N, Zhang W B, Wang J J, Wang C B, Shen Q, Xu Z G, Peng J, Du Y 2022 *Int. J. Refract. Met. Hard Mater.* **104** 105798
- [32] Rahadian H, Bandong S, Widyotriatmo A, Joelianto E 2023 *Alex. Eng. J.* **82** 304
- [33] Zhong L, Guo X, Ding M, Ye Y C, Jiang Y F, Zhu Q, Li J L 2024 *Comput. Electron. Agr.* **217** 108627
- [34] Someh N G, Pishvaei M S, Sadjadi S J, Soltani R 2020 *J. Eval. Clin. Pract.* **26** 1498
- [35] Cervantes J, Lamont F G, Mazahua L R, Lopez A 2020 *Neurocomputing* **408** 189
- [36] Tsai C Y, Kim J, Jin F, Jun M, Cheong M, Yammarino F J 2022 *Leadership Quart.* **33** 101592
- [37] Khakurel H, Tanfique M F N, Roy A, Balasubramanian G, Ouyang G, Cui J, Johnson D D, Devanathan R 2021 *Sci. Rep.* **11** 17149
- [38] Genuer R, Poggi J M, Malot C T, Vialaneix N V 2017 *Big Data Res.* **9** 28

## Hardness prediction of WC-Co cemented carbide based on machine learning model\*

Song Rui    Liu Xue-Mei<sup>†</sup>    Wang Hai-Bin    Lü Hao    Song Xiao-Yan<sup>‡</sup>

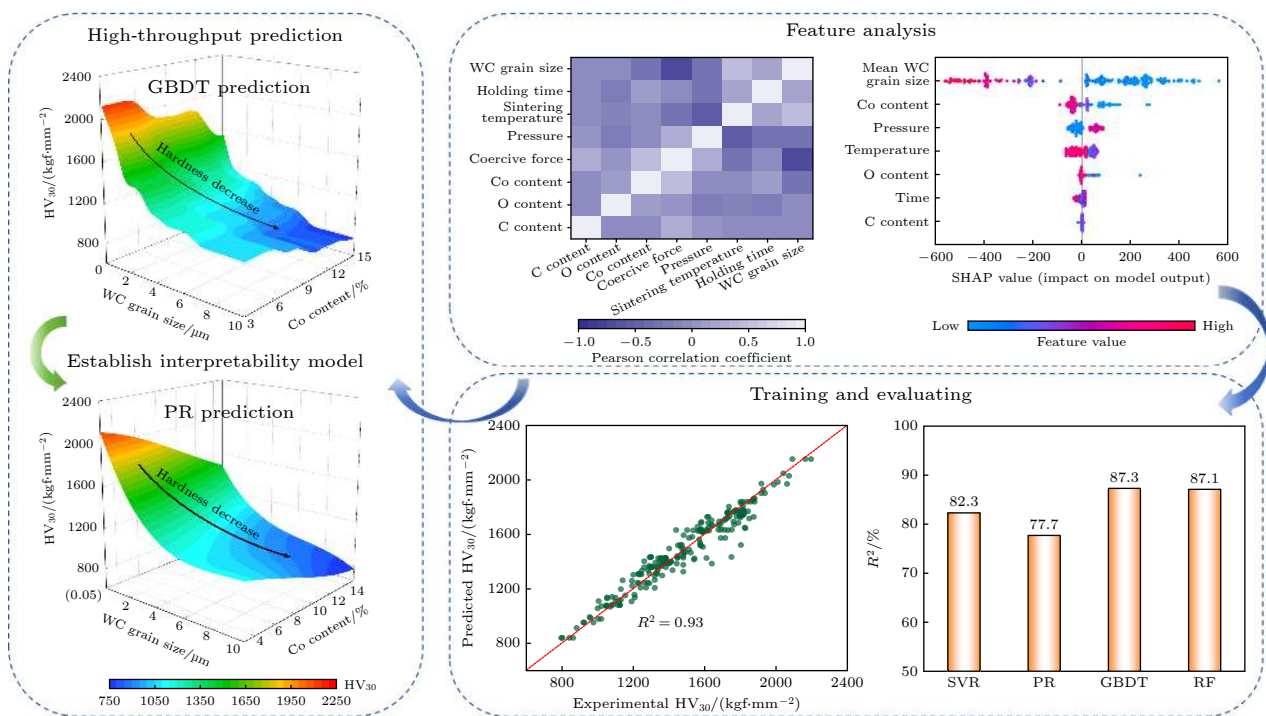
(Key Laboratory of Advanced Functional Materials, Education Ministry of China, College of Materials Science and Engineering, Beijing University of Technology, Beijing 100124, China)

(Received 22 February 2024; revised manuscript received 15 April 2024)

### Abstract

The hardness of cemented carbides is a fundamental property that plays a significant role in their design, preparation, and application evaluation. This study aims to identify the critical factors affecting the hardness of WC-Co cemented carbides and develop a high-throughput predictive model for hardness. A dataset consisting of raw material composition, sintering parameters and characterization results of cemented carbides is constructed

in which the hardness of cemented carbide is set as the target variable. By analyzing the Pearson correlation coefficient, Shapley additive explanations (SHAP) results, WC grain size and Co content are determined to be the key characteristics influencing the hardness of cemented carbide. Subsequently, machine learning models such as support vector regression (SVR), polynomial regression (PR), gradient boosting decision tree (GBDT), and random forest (RF) are optimized to construct prediction models for hardness. Evaluations using 10-fold cross-validation demonstrate that the GBDT algorithm model exhibits the highest accuracy and strong generalization capability, making it most suitable for predicting and analyzing the hardness of cemented carbides. Based on predictions from GBDT algorithm model, PR algorithm model is established to achieve high-precision interpretable prediction of the hardness of cemented carbides. As a result, a quantitative relationship between hardness and Co content and WC grain size is obtained, demonstrating that reducing grain size and Co content is the key to obtaining high hardness of cemented carbide. This research provides a data-driven method for accurately and efficiently predicting cemented carbide properties, presenting valuable insights for the design and development of high-performance cemented carbide materials.



**Keywords:** cemented carbide, hardness, machine learning, high-throughput prediction

**PACS:** 62.20.Qp, 62.25.-g, 74.70.Ad

**DOI:** 10.7498/aps.73.20240284

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 52271085, 92163107, 52171061).

† Corresponding author. E-mail: liuxumei@bjut.edu.cn

‡ Corresponding author. E-mail: xysong@bjut.edu.cn

## 机器学习辅助的WC-Co硬质合金硬度预测

宋睿 刘雪梅 王海滨 吕皓 宋晓艳

### Hardness prediction of WC-Co cemented carbide based on machine learning model

Song Rui Liu Xue-Mei Wang Hai-Bin Lü Hao Song Xiao-Yan

引用信息 Citation: *Acta Physica Sinica*, 73, 126201 (2024) DOI: 10.7498/aps.73.20240284

在线阅读 View online: <https://doi.org/10.7498/aps.73.20240284>

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

#### 机器学习结合固溶强化模型预测高熵合金硬度

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

物理学报. 2023, 72(18): 180701 <https://doi.org/10.7498/aps.72.20230646>

#### WC-Co硬质合金表面石墨烯沉积生长分子动力学仿真研究

Molecular dynamics simulation of graphene deposition and growth on WC-Co cemented carbides

物理学报. 2023, 72(6): 068703 <https://doi.org/10.7498/aps.72.20221332>

#### 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>

#### 蛋白质计算中的机器学习

Machine learning for *in silico* protein research

物理学报. 2024, 73(6): 069301 <https://doi.org/10.7498/aps.73.20231618>

#### 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>