第 49 卷 第 9 期/2022 年 5 月/中国激光 研究论文

中国激光

基于激光诱导击穿光谱与随机森林识别 GCr15 钢的硬度

李 铸[1,2]，张庆永[1]，孔令华[1,2*]，练国富[1]，李 鹏[1,2]

[1]福建工程学院机械与汽车工程学院，福建 福州 350118；

[2]数字福建工业制造物联网实验室，福建 福州 350118

摘要 通过热处理工艺制备了 6 种不同回火温度的 GCr15 钢样品，通过对样品的光谱特性进行分析，发现谱线强度的比值与样品的硬度之间存在线性相关性，并且这种相关性的大小与所选择的分析谱线有关。因此，提出将激光诱导击穿光谱与随机森林算法相结合的方法，对样品的硬度进行研究。利用两种特征选择方法建立了不同的随机森林模型，结果表明，当基于主成分分析选择特征时，所建立的随机森林模型不能对样品的硬度进行正确的识别，而当基于变量重要性提取特征时，所建立的随机森林模型能有效识别样品硬度，并且调节随机森林模型的参数可以使模型的性能得到进一步提高。研究结果表明，激光诱导击穿光谱与随机森林算法相结合是一种新颖的硬度测量技术，可以应用于工程中钢的性能评估。

关键词 光谱学；GCr15 钢；激光诱导击穿光谱；光谱分析；随机森林

中图分类号 0433 **文献标志码** A **DOI**：10.3788/CJL202249.0911002

1 引 言

轴承是现代机械设备中的重要零部件，使用的工况条件比较复杂，对轴承的力学性能有着极高的要求。硬度作为轴承的重要性能之一，在其生产过程中需要对硬度进行检测以满足要求。传统的硬度检测大多属于有损检测，需要破坏原有组件结构，同时采用机械手段在待检测部位切割取样进行实验室分析，这种检测方法不仅速度慢、工作量大，还会对样品造成损伤。因此，探索新的无损检测技术对热处理后的样品硬度进行检测，能够有效避免检测过程对样件的损伤。

激光诱导击穿光谱（Laser-Induced Breakdown Spectroscopy，LIBS）技术是一种原子发射光谱技术，激光作用于样品表面对其进行烧蚀，激发出等离子体，激光作用结束后，等离子体逐渐冷却，这时会辐射出大量的特征谱线，通过对特征谱线的波长和强度进行分析，可以对元素进行识别并对元素含量进行测量。LIBS 技术具有分析速度快，几乎无需样品处理，对样品损伤小等优势，被广泛应用于环境监测[1-2]、材料特性研究[3-4]、煤质分析[5-6]、地质研究[7-8]、太空探测[9]、生物医学研究[10-11]、文物鉴定[12]等领域。在激光诱导击穿光谱测量分析中，激光和物质之间相互作用时等离子体特性会受到待测样品物理化学性能的影响，这种影响会导致发射谱线产生差异，根据材料光谱的这种差异可以对材料的特性进行分析。

Tsuyuki 等[13]利用 LIBS 技术研究了不同抗压强度的混凝土样品，发现可以用谱线强度的比值 $I_{\mathrm{Ca\,II}}/I_{\mathrm{Ca\,I}}$ 表征样品硬度。Aberkane 等[14]利用 LIBS 技术对 Fe-V-C 合金表面硬度与激光诱导等离子体温度的相关性进行了研究，发现激光诱导等离子体温度与维氏表面硬度呈线性关系，谱线强度比 $I_{\mathrm{V\,II}}/I_{\mathrm{V\,I}}$ 与样品表面硬度之间也有良好的正相关关系。Li 等[15]利用 LIBS 技术对不同老化等级的 T91 合金钢进行了研究，实验结果表明，谱线强度的比值（$I_{\mathrm{Cr\,I}}/I_{\mathrm{Fe\,I}}$、$I_{\mathrm{Mo\,I}}/I_{\mathrm{Fe\,I}}$）与老化等级有良好的线性相关性，并且谱线强度比与样品表面硬度之

收稿日期：2021-09-13；**修回日期**：2021-09-26；**录用日期**：2021-10-08

基金项目：福建省科技重大专项专题项目（2020HZ03018）、福建省教育科研专项（GY-Z21004）

通信作者：*15392030898@163.com

间也有较好的关联性。Yang 等[16]利用 LIBS 技术研究了 3D 打印件的硬度，发现离子线与原子线的强度比值与样品硬度之间存在良好的线性关系，利用微量元素的弱自吸收光谱可以获得更好的线性关系。虽然这些方法可以对样品的硬度进行表征，但是还存在谱线选择困难、精度较低等不足之处。因此，近年来，随着机器学习的发展，将算法与 LIBS 技术相结合可以弥补传统 LIBS 分析方法的不足。Huang 等[17]将 LIBS 技术与典型相关分析（CCA）和支持向量回归（SVR）相结合建立了硬度多元定标模型以确定钢的力学性能。Chatterjee 等[18]将 LIBS 与主成分分析（Principal Component Analysis，PCA）法相结合以实现地热土壤样品的快速分类。Zhang 等[19]使用 LIBS 技术并结合偏最小二乘判别分析（PLS-DA），对炉渣样品进行了分类。Dastjerdi 等[20]将 LIBS 技术与支持向量机（SVM）结合，对回收过程中的聚氯乙烯（PVC）聚合物进行了识别和分类。Cui 等[21]将 LIBS 技术与人工神经网络（ANN）相结合，对木材进行了分类。Li 等[22]将 LIBS 技术与 k 最近邻（kNN）分类器相结合，实现了对脂肪、皮肤和肌肉组织的识别。Vrábel 等[23]利用 LIBS 技术并结合多变量数据分析算法（MVDA），实现了对选择性激光熔化（SLM）材料的分类。

随机森林算法（Random Forest，RF）也是机器学习领域中最重要的分类算法之一。它是一种基于多分类器的分类算法，与传统的分类算法相比，具有计算速度快、泛化能力强、可避免过拟合风险等优点。RF 是一种很好的分类算法，已经被证明可以提高分类精度，获得更好的分类效果。已有一些将 LIBS 与 RF 结合使用的研究报道，应用范围包括矿渣分类[24]、红酒产地识别[25]、考古陶瓷分类[26]、铁矿石分类[27]、铝合金分类[28]、钢样分类[29]等。LIBS 结合机器学习方法在物质分类上的应用比较多，所需分类的样品往往都含有不同的元素成分且含量不同，而在元素浓度相同但硬度不同的样品上的应用较少。因此，本文将 RF 与 LIBS 相结合对材料的硬度进行识别，采集了不同硬度的 GCr15 钢样品的 LIBS 数据，构建了随机森林识别模型，研究了 LIBS-RF 方法对 GCr15 钢硬度进行识别的可行性。

2 实验与方法

2.1 样品制备

实验材料选用 GCr15 轴承钢，化学成分如表 1 所示。将原材料升温至 860 ℃后保温 1 h，取出油冷淬火，淬火后马上再进行回火，回火温度分别设置为 150，250，350，450，550，650 ℃，各保温 2 h，然后出炉空冷至室温，得到不同回火温度下的样品 S1（150 ℃）、S2（250 ℃）、S3（350 ℃）、S4（450 ℃）、S5（550 ℃）、和 S6（650 ℃）。将热处理完成后的样品切割成尺寸大小为 12 mm×10 mm×5 mm 的块状，用砂纸打磨后，利用维氏硬度计测量出每种样品的硬度值。材料的热处理工艺参数及硬度如表 2 所示。

表 1 GCr15 钢化学成分

Table 1 Chemical compositions of GCr15 steel

Element	Fe	Cr	C	Si	Mn	P	Mo
Mass fraction /%	Bal.	1.47	0.97	0.24	0.31	0.16	≤0.10

表 2 GCr15 钢的热处理工艺参数及硬度

Table 2 Heat-treatment process parameters and hardness of GCr15 steel

Sample No.	Heat-treatment process parameter		Hardness/ HV
	Oil quenching	Tempering	
S1	860 ℃+1 h	150 ℃+2 h	730.82
S2		250 ℃+2 h	633.56
S3		350 ℃+2 h	547.94
S4		450 ℃+2 h	479.78
S5		550 ℃+2 h	419.66
S6		650 ℃+2 h	292.26

2.2 LIBS 实验

实验所搭建的 LIBS 装置如图 1 所示。激光光源为 Nd:YAG 激光器，工作波长为 1064 nm，脉冲宽度为 9 ns，重复频率为 3 Hz、脉冲能量为 100 mJ；光纤光谱仪的波长范围为 228.0～748.0 nm，平均分辨率为 0.1 nm，积分时间为 1.05 ms。高能脉冲激光通过聚焦的石英凸透镜(focus lens 1)聚焦在样品表面上，样品被激发形成等离子体，等离子体在退激过程中辐射出的特征光谱经聚焦透镜(focus lens 2)聚集后由光纤传输至光谱仪，经光谱仪处理后，光谱数据被传输至电脑。实验前用砂纸对样品表面进行打磨，以免表面杂质对实验结果造成影响。为了减弱样品表面空气击穿的影响，得到稳定的等离子体，将样品放置于三维电控移动平台上，调整样品表面的位置，使其位于合适的位置上，在最佳信噪比延迟时间为 2.8 μs 的条件下采集光谱数据。对于 6 个不同硬度的样品，在每个样品表面以每个点 300 发激光脉冲扫描出 10×10 的点阵，每个样品可得到 30000 个原始光谱。相邻点之间的距离设置为 0.6 mm，以避免消融点之间的影响。

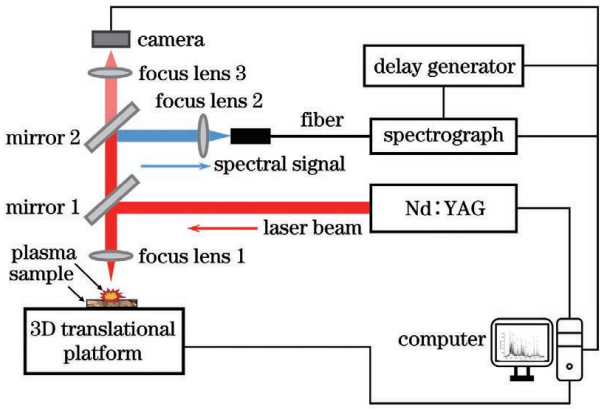

图 1 LIBS 实验装置示意图

Fig. 1 Schematic of LIBS experimental system

2.3 随机森林

RF[30]是一种基于分类与回归树(CART)的高级集成学习算法，其示意图如图 2 所示。在训练时，通过自助法重采样技术从原始数据集中连续生成训练集和测试集，然后采用随机森林将训练集分成多个决策树，最终预测结果由每一个决策树进行投票决定。随机森林中的“随机”主要是指 RF 模型建立过程的两个随机化过程。一是在原始训练数据中随机地有放回地选取数据作为训练样本，这样 RF 模型在面对输入数据的细微变化时变得更加鲁棒，从而具有更好的分类稳定性。另一个是随机森林中的子树的每一个分裂过程并未用到所有的待选特征，而是从所有的待选特征中随机选取一定的特征，之后再在随机选取的特征中选取最优的特征，这样能够使得随机森林中的决策树都能够彼此不同，提升系统的多样性，从而提升分类性能。由于采用的是重采样技术，一些样本可能在同一个训练集中出现多次，而其他一些却可能被忽略，因此，在每一轮采样过程中，训练集中大约包含有 2/3 的原始数据参与建模，称为袋内(in bag)数据，其余 1/3 的训练数据没有参与建模，被称为袋外(out of bag，OOB)数据，OOB 数据可以取代测试集误差估计方法，用于模型的验证。

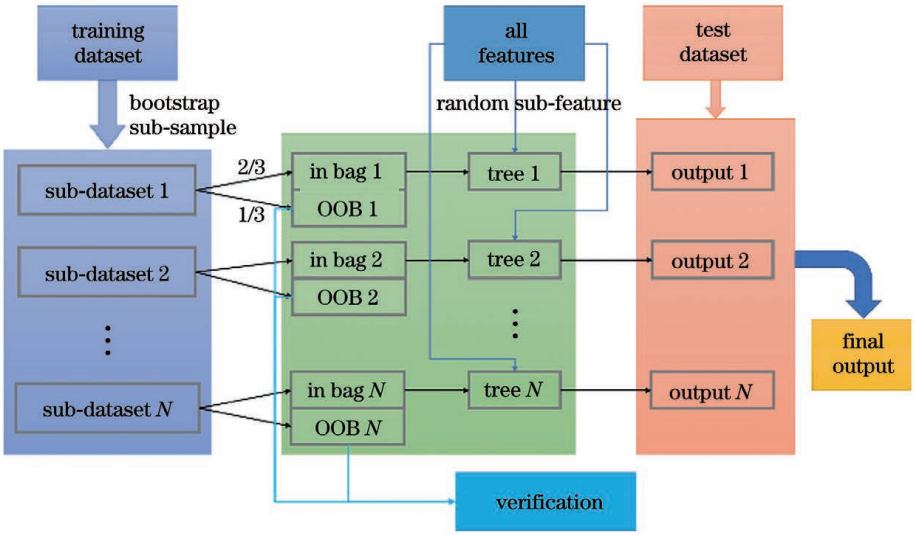

图 2 RF 算法模型示意图

Fig. 2 Schematic of RF algorithm model

3 结果与讨论

3.1 样品的光谱特性

图 3 是 6 种样品的激光诱导击穿光谱图(280～750 nm)。为了研究谱线强度随激光击打次数的变化趋势,选取了 Fe 和 Cr 元素的谱线进行分析,其中Ⅰ表示原子谱线,Ⅱ表示离子谱线。图 4 与图 5 分别为样品 S1 与 S6 中元素的谱线强度随脉冲次数的变化趋势。可以看出,随着激光脉冲次数的增加,前期谱线强度有明显的变化,之后谱线强度逐渐趋于稳定。这是烧蚀坑效应导致的,当激光与样品相互作用时,样品发生烧蚀,烧蚀深度随着脉冲次数的增加而改变,谱线强度会受到一定程度的影响,在早期,烧蚀坑未完全成型,激光烧蚀造成样品表面凹凸不平,所以谱线强度波动较大;随着脉冲次数的继续增加,烧蚀坑成型后,谱线强度相对稳定[31]。因此,可认为烧蚀坑效应对谱线强度的影响不明显,达到一定的击打次数后,不同硬度轴承钢的谱线强度与击打次数无关,因此剔除掉谱线强度波动较大的数据,选择后 100 发(201～300)脉冲的光谱数据作为后续分析。

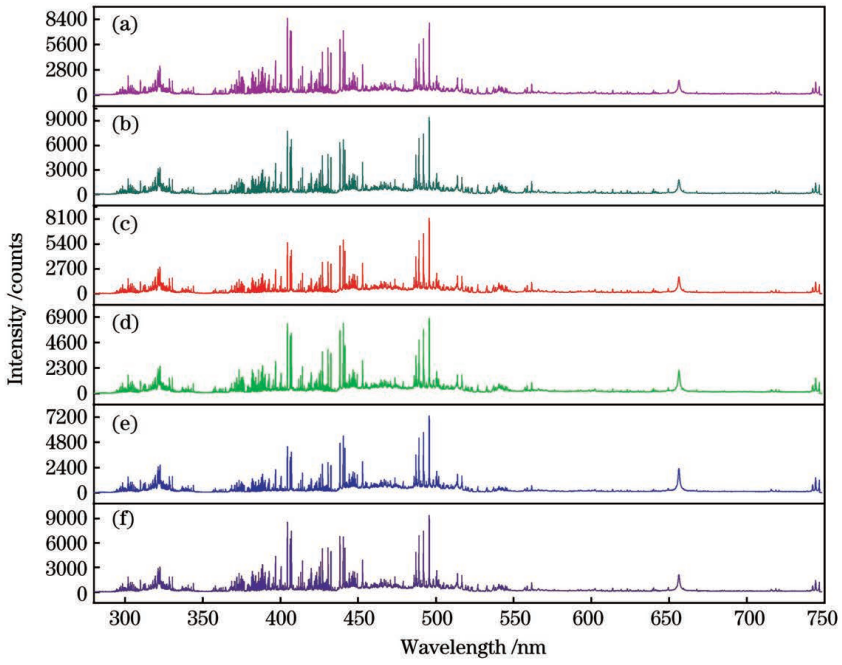

图 3 不同样品的激光诱导击穿光谱图。(a)S1;(b)S2;(c)S3;(d)S4;(e)S5;(f)S6

Fig. 3 LIBS images of different samples. (a) S1; (b) S2; (c) S3; (d) S4; (e) S5; (f) S6

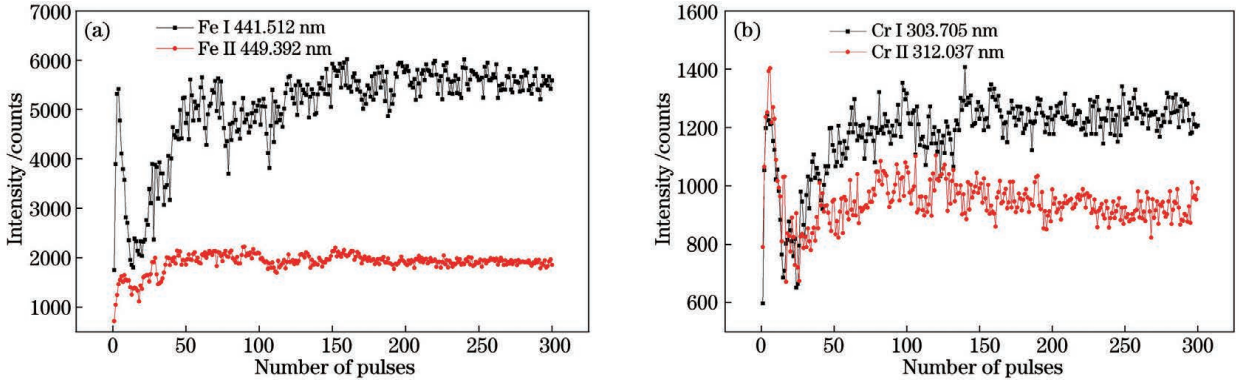

图 4 样品 S1 中元素的谱线强度随脉冲次数的变化趋势。(a)Fe;(b)Cr

Fig. 4 Variation trend of spectral line intensity of elements in sample S1 with pulse number. (a) Fe; (b) Cr

许多研究[13-16]表明,离子与原子谱线强度的比值能够表征样品的硬度。选择了谱线基体元素(Fe Ⅰ 375.745 nm,Fe Ⅱ 316.786 nm)与合金元素(Cr Ⅰ 302.067 nm,Cr Ⅱ 482.413 nm)的谱线,谱线强度的比值与样品的硬度之间的关系如图 6 所示,其中 R^2 表示相关系数。可以发现,

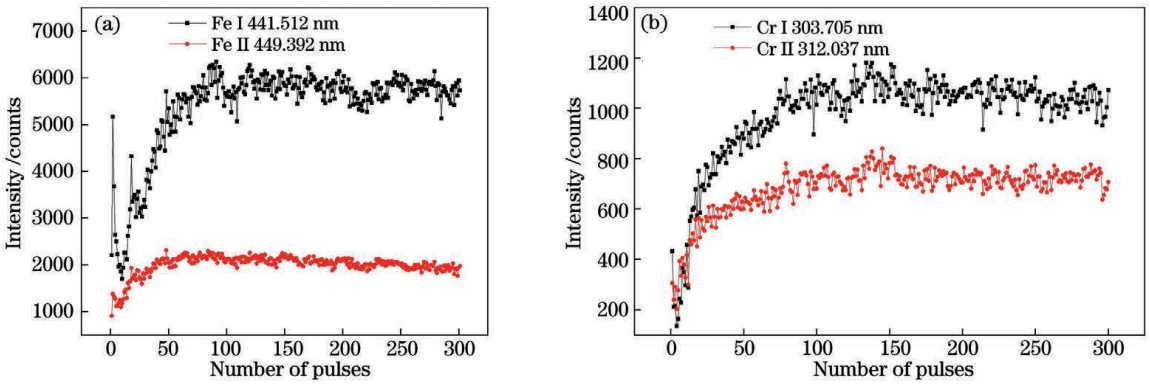

图 5　样品 S6 中元素的谱线强度随脉冲次数的变化趋势。(a)Fe;(b)Cr

Fig. 5　Variation trend of spectral line intensity of elements in sample S6 with pulse number. (a) Fe; (b) Cr

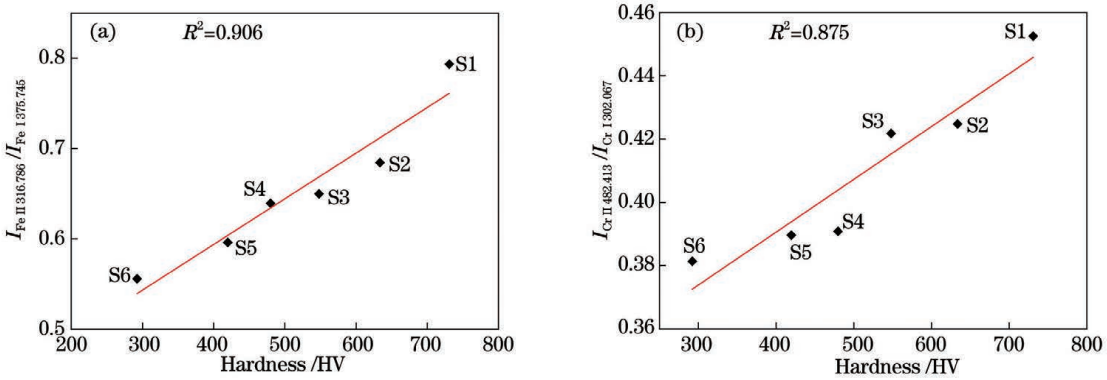

图 6　离子与原子谱线强度的比值与硬度相关性。(a)$I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$;(b)$I_{\mathrm{Cr\ II\ 482.413}}/I_{\mathrm{Cr\ I\ 302.067}}$

Fig. 6　Correlation between ratio of spectral line intensity of ion to that of atom and hardness. (a) $I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$; (b) $I_{\mathrm{Cr\ II\ 482.413}}/I_{\mathrm{Cr\ I\ 302.067}}$

$I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$ 和 $I_{\mathrm{Cr\ II\ 482.413}}/I_{\mathrm{Cr\ I\ 302.067}}$ 与样品硬度之间都存在一定的线性相关关系，并且 $I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$ 与样品硬度之间的相关性更高。这主要是因为样品中各元素含量不一样，各元素在样品中的分布状态也不一样，且激光在与物质相互作用时，不同元素谱线的激发特性存在差异。各样品中 Fe 元素相比其他元素，含量最高，分布较均匀，并且谱线激发特性良好，因此具有较高的相关系数。当利用谱线强度比测量材料的硬度时，只要合理选择谱线，就可以获得较高的相关系数，从而提高硬度的测量精度。

然而，谱线强度的比值虽然可以对样品的硬度进行表征，但是其准确率对谱线的选择有一定的依赖性，这就需要选择合适的谱线进行分析。而随机森林算法作为一种集成学习方法，能够高效地提取和概括光谱特征，本文将激光诱导击穿光谱结合随机森林算法(LIBS-RF)以估计钢样本的硬度。

3.2　随机森林预测准确率

主成分分析法[32]是一种对高维数据降维的统计方法，它将原来变量重新组合成一组新的相互无关的综合变量，根据实际需要从中取出几个较少的综合变量以尽可能多地反映原来变量的信息。LIBS 数据集包含了大量反映样品特征的信息，因此可以通过 PCA 方法对原始光谱数据进行降维处理，提取光谱特征，降低数据复杂度。每种样品采集了 100 组光谱数据，对这 6 种样品的 600 个样本数据进行 PCA 处理。图 7 为提取到的分值最高的 15 个主成分，可以看出，随着提取的主成分数量的增加，

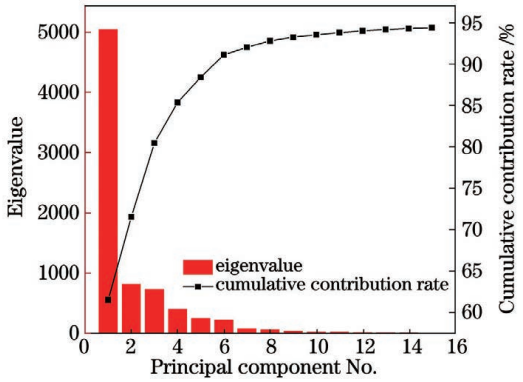

图 7　前 15 个主成分的碎石图

Fig. 7　Gravel plots of first 15 principal components

累计贡献率逐渐上升，15 个主成分的累计贡献率达到了 95%，即保留了原始数据的 95%的信息。利用提取到的 15 个主成分建立随机森林硬度预测模型，其结果如图 8 所示。可以发现，在经过 PCA 降维处理后再建立随机森林模型，预测准确率最高的为样品 S2，为 69%（20/29），其次是样品 S3，为 52%（12/23），其余样品的预测准确率均较低，样品 S5 和 S6 的预测准确率几乎为零。经过分析发现，一方面，PCA 属于一种无监督学习算法，它将所有的样本作为一个整体，寻找均方误差最小条件下的从高维空间到低维空间的最优线性映射，然而对于本文的数据集，可能需要非线性的映射才能找到恰当的低维嵌入；另一方面，降维可以降低程序的计算复杂度，代价是丢弃了原始数据的一些信息，比如忽略了类别属性，而所忽略的投影方向有可能刚好包含了重要的分类信息。因此，主成分分析法不能有效进行预测。

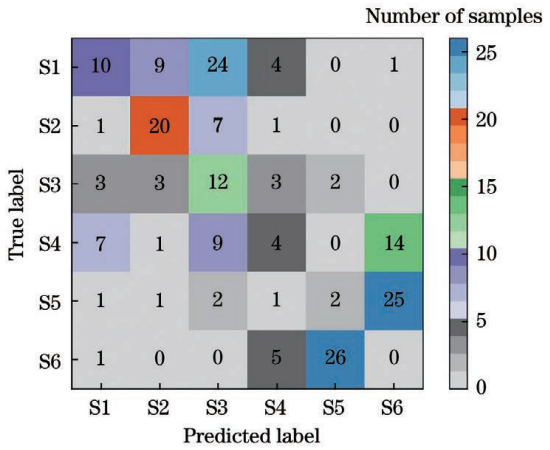

图 8　RF 的预测结果

Fig. 8　Predicted result of RF

PCA 降维处理后建立的随机森林模型不能有效地进行预测，尝试将全谱数据用于构建随机森林模型，其结果如表 3 所示。可以发现，模型能够较为准确地进行预测。但是，在实际应用中，光谱数据往往具有相当大的维度，如果将全部特征用于模型的构建，会增加算法的复杂性，效率也会降低。因此，基于特征的重要性进行数据降维，进而提取有效光谱特征。取到的主要光谱特征如表 4 所示，将其作为随机森林的输入进行预测，其结果如表 5 所示。与将全谱数据用于模型建立相比，准确率有较为明显的提高。这是因为全谱数据的信息量虽然最为全面，但是也包含了大量的冗余信息，这些冗余信息参与模型的训练会造成模型性能的降低，从而使模型的精度下降。

表 3　基于全谱数据的 RF 模型的预测结果

Table 3　Prediction results of RF model based on full-spectrum data

Sample No.	Classification accuracy /%
S1	0.833(40/48)
S2	1(29/29)
S3	0.870(20/23)
S4	0.771(27/35)
S5	1(31/31)
S6	1(32/32)
Average	0.904

表 4　用于 RF 模型训练的特征谱线

Table 4　Characteristic spectral lines used for RF model training

Line type	Wavelength /nm
Fe Ⅰ	300.303 304.760 310.030 315.331 315.704 358.119 364.742 367.687 375.745 511.034
Fe Ⅱ	296.994 302.547 316.786 325.877 327.349 450.828 492.392 510.061 510.745 515.991
Cr Ⅰ	302.067 303.704 451.543 454.070 455.617 458.390 460.010 465.472 469.894 483.165 519.199 558.805
Cr Ⅱ	482.413
Mn Ⅰ	382.351，476.586，482.352
Mn Ⅱ	299.361

表 5　基于特征谱线的 RF 模型的预测结果

Table 5　Prediction results of RF model based on characteristic spectral lines

Sample No.	Classification accuracy /%
S1	0.979(47/48)
S2	0.931(27/29)
S3	0.957(22/23)
S4	0.971(34/35)
S5	0.968(30/31)
S6	0.938(30/32)
Average	0.960

3.3　随机森林模型参数对模型性能的影响

决策树的数量是随机森林算法中的关键参数。根据随机森林算法的理论，在处理分类问题时，该模型中的决策树的数量很有意义。图 9 所示为 OOB

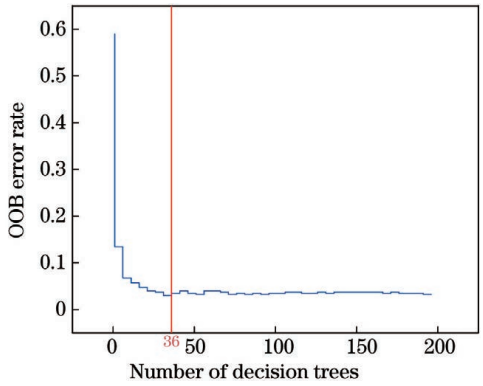

图 9 决策树数量对模型性能的影响

Fig. 9 Effect of number of decision trees on model performance

错误率与决策树的关系，可以看出，当决策树的数量从 1 增大至 20 时，模型的 OOB 错误率迅速减小。这也符合随机森林的特点，在一定范围内，决策树的数量越多，模型效果越好。然而，结果也表明，当树的数量达到一定程度时，OOB 错误率在一定水平范围内波动，当树的数量为 36 时，模型获得最低的 OOB 错误率。使用 OOB 误差估计对模型性能进行验证，OOB 误差越低，模型效果越好。除此之外，OOB 错误率还可以评估其他参数对分类模型的影响，理论上，分类树的预测误差有一个上限，当 OOB 误差达到最小值时，该选择被认为是基于 OOB 误差的最优选择。因此，在验证随机森林模型性能时，可以直接使用 OOB 误差估计而不必使用其他方法，并且 OOB 误差估计不需要大量计算，在一定程度上也提高了模型的效率。随机特征数目作为随机森林另一个关键参数，也会对模型的精度产生影响。图 10 为 OOB 错误率与随机特征数目的关系，可以看出，随着特征数目的增加，OOB 错误率呈现出先下降然后上升并逐渐趋于平稳的趋势，当候选特征

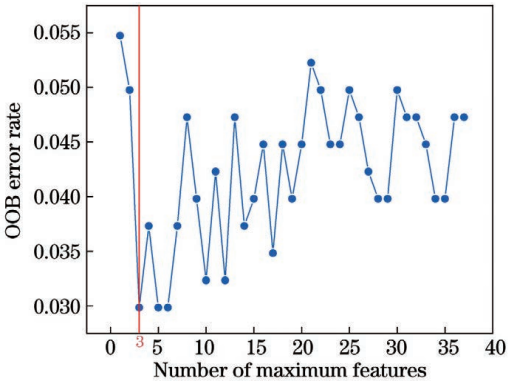

图 10 最大特征数量对模型性能的影响

Fig. 10 Effect of maximum number of features on model performance

数量为 3 时，获得最低的 OOB 错误率，模型的性能最好。这也说明了随机森林算法在处理高维数据时，并不需要全部特征就能获得良好的性能，模型的复杂度低，模型的效率得到提高。

4 结 论

利用激光诱导击穿光谱和随机森林模型对不同硬度的 GCr15 钢样品进行了研究。实验结果表明，基体元素与合金元素的谱线强度比值与样品的硬度都呈现出线性相关性，并且 $I_{\mathrm{Fe\,II\ 316.786}}/I_{\mathrm{Fe\,I\ 375.745}}$ 与硬度的线性相关性更高，说明基于谱线强度比值的方法对谱线的选择具有一定的依赖性。提出了 LIBS-RF 方法来估计样品的硬度。首先利用 PCA 对原始数据进行降维处理，再建立随机森林模型，发现该方法不能对样品的硬度进行有效的预测。然后基于变量的重要性选择特征光谱，进而建立随机森林模型，结果显示，基于全谱数据的 LIBS-RF 模型的预测精度低于基于部分特征光谱数据的 LIBS-RF 模型，这是因为全谱数据包含了大量噪声和其他冗余信息，这些冗余信息也参与模型的训练，造成了模型精度的下降。此外还发现，随着决策树数量和随机特征数量的增加，模型的预测精度上升到一定值后保持相对稳定。因此，可以对模型的参数进行合理的调整，在满足精度要求的同时，降低算法的复杂度，提高算法的效率。LIBS-RF 作为一种新颖的测量硬度的技术，相较于传统的 LIBS 硬度测量方法，具有更加简单、快捷的优点，为工程实践应用提供了理论依据。

参 考 文 献

[1] Kumar R, Devanathan A, Mishra N L, et al. Quantification of heavy metal contamination in soil and plants near a leather tanning industrial area using libs and TXRF[J]. Journal of Applied Spectroscopy, 2019, 86(5): 942-947.

[2] Qu Y F, Zhang Q H, Yin W Y, et al. Real-time *in situ* detection of the local air pollution with laser-induced breakdown spectroscopy[J]. Optics Express, 2019, 27(12): A790-A799.

[3] Imran M, Sun L Y, Liu P, et al. Depth profiling of tungsten coating layer on CuCrZr alloy using LIBS approach[J]. Surface and Interface Analysis, 2019, 51(2): 210-218.

[4] 雷鹏达, 付洪波, 易定容, 等. 激光诱导击穿光谱表征熔覆层缺陷程度[J]. 中国激光, 2020, 47(4): 0411001.

Lei P D, Fu H B, Yi D R, et al. Characterization of cladding defects via laser-induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2020, 47(4): 0411001.

[5] Yuan T B, Wang Z, Lui S L, et al. Coal property analysis using laser-induced breakdown spectroscopy [J]. Journal of Analytical Atomic Spectrometry, 2013, 28(7): 1045-1053.

[6] Hou Z Y, Wang Z, Yuan T B, et al. A hybrid quantification model and its application for coal analysis using laser induced breakdown spectroscopy [J]. Journal of Analytical Atomic Spectrometry, 2016, 31(3): 722-736.

[7] Wang C, Wang J, Wang J, et al. Classification of 13 original rock samples by laser induced breakdown spectroscopy[J]. Laser Physics, 2021, 31(3): 035601.

[8] 邱苏玲, 李安, 王宪双, 等. 基于激光诱导击穿光谱的矿石中铁含量的高准确度定量分析[J]. 中国激光, 2021, 48(16): 1611002.
Qiu S L, Li A, Wang X S, et al. High-accuracy quantitatively analysis of iron content in mineral based on laser-induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2021, 48(16): 1611002.

[9] Rauschenbach I, Jessberger E K, Pavlov S G, et al. Miniaturized laser-induced breakdown spectroscopy for the *in situ* analysis of the Martian surface: calibration and quantification [J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2010, 65(8): 758-768.

[10] Wang Q Q, Xiangli W T, Teng G E, et al. A brief review of laser-induced breakdown spectroscopy for human and animal soft tissues: pathological diagnosis and physiological detection[J]. Applied Spectroscopy Reviews, 2021, 56(3): 221-241.

[11] 刘燕德, 高雪, 程梦杰, 等. 激光诱导击穿光谱技术检测油茶炭疽病[J]. 激光与光电子学进展, 2020, 57(9): 093006.
Liu Y D, Gao X, Cheng M J, et al. Detection of anthracnose in camellia oleifera based on laser-induced breakdown spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(9): 093006.

[12] Yin Y P, Yu Z R, Sun D X, et al. A potential method to determine pigment particle size on ancient murals using laser induced breakdown spectroscopy and chemometric analysis[J]. Analytical Methods, 2021, 13(11): 1381-1391.

[13] Tsuyuki K, Miura S, Idris N, et al. Measurement of concrete strength using the emission intensity ratio between Ca(II) 396.8 nm and Ca(I) 422.6 nm in a Nd∶YAG laser-induced plasma[J]. Applied Spectroscopy, 2006, 60(1): 61-64.

[14] Aberkane S M, Bendib A, Yahiaoui K, et al. Correlation between Fe-V-C alloys surface hardness and plasma temperature via LIBS technique [J]. Applied Surface Science, 2014, 301: 225-229.

[15] Li J, Lu J D, Dai Y, et al. Correlation between aging grade of T91 steel and spectral characteristics of the laser-induced plasma[J]. Applied Surface Science, 2015, 346: 302-310.

[16] Yang J W, Kong L H, Lian G F, et al. Surface hardness determination of 3D printed parts using laser-induced breakdown spectroscopy[J]. Applied Optics, 2021, 60(3): 499-504.

[17] Huang J W, Dong M R, Lu S Z, et al. Estimation of the mechanical properties of steel via LIBS combined with canonical correlation analysis (CCA) and support vector regression (SVR)[J]. Journal of Analytical Atomic Spectrometry, 2018, 33(5): 720-729.

[18] Chatterjee S, Singh M, Biswal B P, et al. Application of laser-induced breakdown spectroscopy (LIBS) coupled with PCA for rapid classification of soil samples in geothermal areas[J]. Analytical and Bioanalytical Chemistry, 2019, 411(13): 2855-2866.

[19] Zhang T L, Wu S, Dong J, et al. Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (LIBS) coupled with support vector machine (SVM) and partial least square (PLS) methods[J]. Journal of Analytical Atomic Spectrometry, 2015, 30(2): 368-374.

[20] Dastjerdi M V, Mousavi S J, Soltanolkotabi M, et al. Identification and sorting of PVC polymer in recycling process by laser-induced breakdown spectroscopy (LIBS) combined with support vector machine (SVM) model[J]. Iranian Journal of Science and Technology, Transactions A: Science, 2018, 42(2): 959-965.

[21] Cui X T, Wang Q Q, Zhao Y, et al. Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN)[J]. Applied Physics B, 2019, 125(4): 1-12.

[22] Li X H, Yang S B, Fan R W, et al. Discrimination of soft tissues using laser-induced breakdown spectroscopy in combination with k nearest neighbors (kNN) and support vector machine (SVM) classifiers [J]. Optics & Laser Technology, 2018, 102: 233-239.

[23] Vrábel J, Pořízka P, Klus J, et al. Classification of materials for selective laser melting by laser-induced breakdown spectroscopy[J]. Chemical Papers, 2019, 73(12): 2897-2905.

[24] Tang H S, Zhang T L, Yang X F, et al.

Classification of different types of slag samples by laser-induced breakdown spectroscopy (LIBS) coupled with random forest based on variable importance (VIRF)[J]. Analytical Methods, 2015, 7(21): 9171-9176.

[25] Tian Y, Yan C H, Zhang T L, et al. Classification of wines according to their production regions with the contained trace elements using laser-induced breakdown spectroscopy[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2017, 135: 91-101.

[26] Ruan F Q, Hou L, Zhang T L, et al. A novel hybrid filter/wrapper method for feature selection in archaeological ceramics classification by laser-induced breakdown spectroscopy[J]. The Analyst, 2021, 146(3): 1023-1031.

[27] Sheng L W, Zhang T L, Niu G H, et al. Classification of iron ores by laser-induced breakdown spectroscopy (LIBS) combined with random forest (RF)[J]. Journal of Analytical Atomic Spectrometry, 2015, 30(2): 453-458.

[28] Zhan L Y, Ma X H, Fang W Q, et al. A rapid classification method of aluminum alloy based on laser-induced breakdown spectroscopy and random forest algorithm[J]. Plasma Science and Technology, 2019, 21(3): 034018.

[29] Zhang T L, Xia D H, Tang H S, et al. Classification of steel samples by laser-induced breakdown spectroscopy and random forest[J]. Chemometrics and Intelligent Laboratory Systems, 2016, 157: 196-201.

[30] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[31] Lu S Z, Dong M R, Huang J W, et al. Estimation of the aging grade of T91 steel by laser-induced breakdown spectroscopy coupled with support vector machines[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 140: 35-43.

[32] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1/2/3): 37-52.

Hardness Characterization of GCr15 Steel Based on Laser-Induced Breakdown Spectroscopy and Random Forest

Li Zhu[1,2], Zhang Qingyong[1], Kong Linghua[1,2*], Lian Guofu[1], Li Peng[1,2]

[1] *School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, Fujian, China;*

[2] *Digital Fujian Industrial Manufacturing IoT Lab, Fuzhou 350118, Fujian, China*

Abstract

Objective Bearing is an important part of modern machinery and equipment, in which the working conditions are much complicated. Therefore, there exist extremely high requirements for the mechanical performance of bearings. As one of the important properties of bearings, hardness needs to be tested to meet these requirements during the production process. The traditional hardness test is a mostly destructive one, which requires the structural destruction of the original component, and the use of mechanical means is done to cut and sample the parts to be tested for the laboratory analysis. This test method is not only slow and heavy, but also damages the samples. The laser-induced breakdown spectroscopy (LIBS) technology, as an atomic emission spectroscopy technology, has the advantages of fast analysis speed, almost no sample processing, and less damage to samples, which is widely used in environmental monitoring, material characteristic research, coal quality analysis, and other fields. The interaction between laser and substance and the plasma characteristics are affected by the physical and chemical material characteristics of the samples to be tested. This effect causes the difference in the characteristics of emission spectra. This difference in the material spectra can affect the characteristics of the material. However, although the spectral line intensity ratio can be used to characterize the hardness of the samples, there are still some shortcomings such as difficulty in spectral line selection and low accuracy. Therefore, in recent years, with the development of machine learning, combining algorithms with LIBS can make up for the shortcomings of the traditional LIBS analysis methods. As an integrated learning method, the random forest (RF) algorithm can efficiently extract and generalize spectral features. This research attempts to use LIBS combined with the RF algorithm (LIBS-RF) to estimate the hardness of steel samples.

Methods In order to measure the hardness of GCr15 steels, a method combining LIBF and RF is proposed to study

the hardness of the samples. First, six GCr15 steel samples with different tempering temperatures are prepared by the heat-treatment process, and an LIBS experimental device is built. A 10 × 10 lattice is scanned with 300 laser pulses per point on each sample surface to obtain the original spectral data. Then the spectral characteristics of the samples are analyzed, and a correlation between the spectral line intensity ratio and the hardness is established. Finally, a RF model is established based on the principal component analysis (PCA) method and the variable importance method.

Results and Discussions The RF model is established based on the PCA method, and the result is shown in Fig. 8. It can be found that the RF model is established after PCA dimensionality reduction. The highest prediction accuracy for sample S2 is 69% (20/29) followed by 52% (12/23) for sample S3, and the rest prediction accuracy rates of the samples are all low. The prediction accuracy rates of samples S5 and S6 are almost zero. RF is used to reduce data dimensionality based on the importance of features and extract the effective spectral features. The prediction results of the model are shown in Table 5. Compared with the model based on the PCA method, the accuracy rate is significantly improved, and its average accuracy reaches 0.96. In addition, through the study on the effects of the number of decision trees (Fig. 9) and the number of random features (Fig. 10) on the performance of the model, it can be found that the number of decision trees and the number of random features are not that the larger the better. With the increase of the number of decision trees and the number of random features, the prediction accuracy of the model first increases to a certain value and then remains relatively stable.

Conclusions In this paper, the laser-induced breakdown spectroscopy and random forest are combined to study the GCr15 steel samples with different hardness. Through the linear relationship between the spectral line intensity ratio ($I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$ and $I_{\mathrm{Cr\ II\ 482.413}}/I_{\mathrm{Cr\ I\ 302.067}}$) and the sample hardness, the experimental results show that the spectral line intensity ratio of the matrix element to the alloy element and the hardness of the sample present a linear correlation, in which the linear correlation between $I_{\mathrm{Fe\ II\ 316.786}}/I_{\mathrm{Fe\ I\ 375.745}}$ and hardness is higher, indicating that the method of the spectral line intensity ratio has a certain dependence on the selection of spectral lines. The LIBS-RF method is proposed to estimate the hardness of the sample. First, PCA is used to reduce the dimensionality of the original data, and subsequently a random forest model is established. It is found that the hardness of the sample can not be effectively predicted. Then, the random forest is used to select feature spectra based on the importance of the variables to establish a random forest model. The results show that the prediction accuracy of the LIBS-RF model based on the full-spectrum data is lower than that based on the partial characteristic spectrum data. This is because the full-spectrum data also contains a lot of noise and other redundant information, which also participates in the training of the model and results in a decrease in model accuracy. In addition, it is found that as the number of decision trees and the number of random features increase, the prediction accuracy of the model increases to a certain value and remains relatively stable. Based on this, the parameters of the model can be adjusted reasonably while satisfying the accuracy requirements, the complexity of the algorithm is reduced, and the efficiency of the algorithm is improved. Above all, as a novel hardness measurement technology, LIBS-RF has the advantages of simpler and faster than the traditional LIBS hardness measurement, and the results here provide a theoretical basis for engineering practice applications.

Key words spectroscopy; GCr15 steel; laser-induced breakdown spectroscopy; spectral analysis; random forest