

光纤传能的移动式激光诱导击穿光谱钢铁快速检测与分类

李文鑫¹, 陈光辉^{1,3}, 曾庆栋^{1,2*}, 袁梦甜^{1,3}, 何武光¹, 江泽方¹,
刘洋¹, 聂长江¹, 余华清¹, 郭连波²

1. 湖北工程学院物理与电子信息工程学院, 湖北 孝感 432000
2. 华中科技大学武汉光电国家研究中心, 湖北 武汉 430074
3. 湖北大学物理与电子科学学院, 湖北 武汉 430062

摘要 为了实现工业现场对特种钢材的快速检测与种类识别, 采用基于光纤传能的移动式激光诱导击穿光谱(LIBS)样机对14种特种钢材进行光谱数据的采集与分析, 采用预选谱线并遍历组合的降维方法与支持向量机(SVM)相结合的算法对特钢材料的光谱进行快速分类。分别将原始光谱数据、归一化处理后的光谱数据、归一化处理+遍历组合优选谱线数据作为SVM分类模型的输入向量, 并对比了不同输入向量下模型对特钢识别的准确度。结果表明: 在事先选出的51条特征谱线作为输入变量的基础上, 归一化光谱数据作为SVM分类模型的输入特征时, 识别准确度达到95.71%, 明显高于使用原始光谱数据作为输入向量时SVM分类模型的准确度11.43%。进一步地, 使用MATLAB程序遍历谱线组合, 通过遍历各种谱线组合选出最优的输入谱线组合, 当优选6条特定的谱线时, 对特钢种类识别的准确度达到100%, 且建模速度也有相应提升。可以看出, 当预选出大量常见特征数据时, 机器自动选取特征与人工挑选谱线相比, 具有明显优势, 基于此降维方法的SVM算法模型在LIBS快速分类技术中具有很好的工业应用前景。

关键词 激光诱导击穿光谱; SVM; 谱线遍历组合; 降维; 钢铁分类

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)08-2638-06

引言

钢铁行业是我国国民经济的支柱性产业, 是关系到国计民生的基础性行业。2018年我国全年钢铁产量已占世界总产量的50%。最近几年由于产能过剩, 中低端和粗钢的生产所占比重很大, 因此大部分钢铁行业都处于亏损状态, 未来加大高端钢材的产量是改变严峻形势的唯一方式。在实际生产中, 根据不同用途需要向钢内加入不同的合金元素来改变钢铁的某方面的性能, 比如硫(S)可以改善钢的切削性、加工性和磁性, 但也会引起钢的热脆性, 降低钢的机械性能, 如使疲劳极限、塑性和耐磨性显著下降等, 影响钢件的使用寿命; 磷(P)具有强烈的固溶强化作用, 可以增加钢的强度和硬度, 但也会降低钢的塑性和韧性等等。由于不能准确识别废材所含材料就无法合理的再利用, 每年有大量的废弃钢铁的堆积, 这不仅污染环境也是资源的浪费。为了提高钢铁废

弃物的回收利用率, 如何对钢铁材料进行快速检测分类成为了一个新的研究热点。现有的元素分析方法有X射线荧光光谱分析方法、原子吸收光谱分析技术和电感耦合等离子体-原子发射光谱分析技术等, 由于这些技术都有各自的缺点和不足^[1-4], 尚不能满足快速检测的需要, 因此急需一种新的快速在线检测技术。

激光诱导击穿光谱(laser induced breakdown spectroscopy, LIBS)是一种近年来发展迅速的原子发射光谱技术, 它采用高能量激光脉冲聚焦到样品表面产生等离子体, 通过对等离子体中原子和离子能级跃迁辐射出的特征光谱采集和分析, 获得被测样品中所含的元素种类及其含量^[5-6]。LIBS技术具有无需进行样品预处理, 分析速度快, 非接触, 对样品几乎无损等优点, 在工业生产实时在线监测方面具有很大的发展潜力^[7-9]。采用LIBS技术结合分类算法对材料进行快速分类是当前研究的一个热点。

传统的分类算法一般采用K临近法, 然而该算法的特征

收稿日期: 2020-08-04, 修订日期: 2020-12-19

基金项目: 国家自然科学基金项目(61705064, 11647122), 湖北省自然科学基金项目(2018CFB773), 湖北省教育厅团队研究项目(T201617)资助

作者简介: 李文鑫, 1999年生, 湖北工程学院物理与电子信息工程学院本科生 e-mail: 3300144190@qq.com

* 通讯作者 e-mail: jerry-z@hbeu.edu.cn

谱线选取较为复杂。由于钢铁合金中的基体元素为 Fe 元素, 其中还含有 Cr, Ni, C, Mn, Ti, Mo, Cu 等多种元素, 相比于其他类型材料, 钢铁合金样品的 LIBS 谱线异常丰富, 各元素间的相互干扰更为复杂, 为 LIBS 定量分析带来了很大挑战^[10-11]。

支持向量机(support vector machine, SVM)是一种基于结构风险最小化准则的学习方法, 它能够利用核函数变换将原始非线性数据转化为高维线性数据, 剔除大量冗余数据, 具有很强的鲁棒性和容错性。2015 年, 杨友盛等^[12]利用 SVM 强大的分类功能, 通过 Si 和 Mn 对应的特征谱线波长和光谱强度, 利用 Si、Mn 含量和温度来判定转炉终点。结果显示该模型的准确率达 98% 以上, 证明了在实验环境不变的情况下, SVM 分类模型可以用在光谱定性分析上。2016 年, 谷艳红等^[13]对土壤中的 Cr 元素进行定量分析, 采用 LIBS 技术结合 SVM 的方法, 相比于传统定量分析方法, 大大提高了定量分析的精度, 其定标曲线拟合相关系数由 0.689 提高到 0.998, 表明 SVM 算法具有良好的实用性。然而, 在使用 SVM 算法时, 当直接输入大量特征时容易发生过拟合, 从而导致模型的泛化程度降低, 此时需要对输入向量进行降维处理。合理地选择样品组成元素的特征谱线组合作为输入, 可降低输入向量的维度。人工选择特征谱线组合耗时繁琐且效果难以保证。采用谱线遍历组合优化输入向量的降维方法可对所有组合测试, 寻找最优的输入向量, 但相关工作鲜有报道。

本工作采用基于光纤传能的移动式 LIBS 系统结合 SVM 算法对特种钢材进行快速检测和分类, 并提出一种基于预选谱线然后遍历组合的方法对输入向量进行优化降维, 建立基于 SVM 算法的光谱识别模型, 实现对不同牌号的特种钢材的快速分类。

1 实验部分

1.1 仪器

基于光纤传能的移动式 LIBS 实验装置如图 1 所示, 采用的激光器为一款紧凑型调 Q 的 Nd:YAG 激光器(型号: Ultra 50, 美国 Bigsky 公司生产), 该激光器输出波长为 532 nm, 脉冲重复频率为 10 Hz。输出的激光脉冲经过耦合模块耦合进一根芯径为 1 mm 的传能光纤, 经光纤输出的最大单脉冲激光能量约为 29 mJ, 经过准直透镜进行准直, 然后被二向色镜反射后, 经过聚焦透镜聚焦在样品靶材表面, 激发产生等离子体光谱。等离子体光谱经透镜采集耦合到光纤, 然后传输到一台紧凑型光纤光谱仪(型号为 Avaspec-2048 USB2, 10 μm 狭缝, 2 400 线 $\cdot\text{mm}^{-1}$ (VE)光栅)。该光谱仪的光谱波长范围约为 290~1 020 nm, 光谱分辨率为 0.08~0.11 nm。配有一个门控 2048 像素的 CCD 阵列探测器(型号为 Sony 554), 主要功能是将光信号转换为电信号。光谱信号随后通过 USB 接口传输并显示在笔记本电脑上。每次光谱采集的积分时间设置为 1.1 ms, 采集延迟时间设为 1.3 μs 。

1.2 样本

实验采用 14 个特钢材料作为分析样品, 分别对其进行

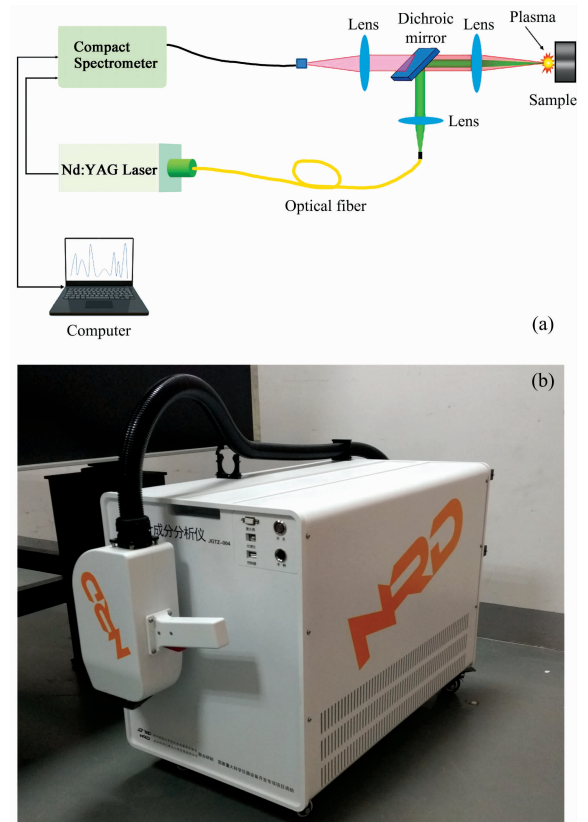


图 1 LIBS 实验装置

(a): 原理框图; (b): 样机

Fig. 1 LIBS system

(a): Schematic; (b): Prototype

编号, 各样品中各元素的参考浓度(Wt%)和对应的编号如表 1 所示(该参考浓度由生产厂家采用火花直读法测得, 分析精度在 5% 以内)。实验参数设置如 1.1 节所述, 对每个样品采集 30 幅光谱, 每幅光谱由 10 个激光脉冲产生的光谱经过平均后得到, 14 个样品共采集 420 幅光谱。

1.3 SVM 算法简介

SVM 算法是在统计学理论上发展起来的一种机器学习算法, 是一种二分类模型, 其基本模型是定义在特征空间上的最大间隔分类器, 在解决小样本、非线性及高维模式识别问题时具有许多特有的优势。本工作使用中国台湾大学林智仁(Lin Chih-Jen)教授等开发设计的一个简单易用和快速有效的 SVM 软件包 libsvm-3.23 进行数据处理。将数据分为两部分: 训练集和测试集, 训练集与测试集设置相同的标签, 每个输入都是一组特征向量与一个判断样品属性的标签值。SVM 工具箱先通过训练集计算出一个分类模型, 然后对测试集进行预测, 并将预测的标签与真实标签做出对比, 以验证模型的准确性。

在等离子体光谱中, 由于激光能量的波动、样品的不均匀性和激光与物质相互作用过程的复杂性, 单一元素的校准模型往往不能满足定量分析的要求, 特别是基体元素为 Fe 时, 谱峰重叠严重, 导致单一元素的特征谱线强度稳定性较差^[15]。因此难以用单一元素的特征谱线作为 SVM 的特征参

表 1 14 个钢铁样品中各元素的含量信息 (Wt%)

Table 1 The concentration information of each element in 14 types of steel samples

编号	C	Mn	Si	P	Cr	Ni	Mo	V	Cu
1	0.09	0.44	0.33	0.007	8.64	0.2	0.9	0.19	0.11
2	0.12	0.50	0.26	0.006	1.10	0.18	0.27	0.2	0.09
3	0.20	0.25	0.29	0.005	2.57	0.2	0.35	0.02	0.07
4	0.16	0.93	0.37	0.007	0.24	1.12	0.31	0.004	0.58
5	0.41	0.54	0.19	0.019	0.94	0.02	0.17	0.006	0.03
6	0.38	0.77	0.24	0.021	1.02	0.02	0.2	0.005	0.02
7	0.11	0.47	0.37	0.005	2.13	0.15	0.94	0	0
8	0.26	0.65	0.22	0.018	0	0	0	0	0
9	0.19	1.57	0.31	0.021	0	0	0	0.002	0
10	0.374	0.372	0.292	0.0053	1.524	1.419	0.251	0	0.074
11	0.422	0.66	0.349	0.021	1.05	0.066	0.195	0.012	0.063
12	0.399	0.365	1	0.013	5.02	0.281	1.14	0.789	0.094
13	0.15	1.05	0.377	0.01	0.117	1.17	0.384	0	0.605
14	0.107	0.333	0.327	0.008	8.22	0.123	0.9	0.236	0.115

数来建立定量分析模型去准确识别钢铁种类, 而采用多元素的多条谱线信息输入支持向量机模型时, 模型训练效果较好, 主要是因为多种谱线信息的输入可以有效校正基体效应的影响。

2 结果与讨论

2.1 特征光谱的选取

以美国国家标准与技术研究院 (National Institute of Standards and Technology, NIST) 的原子光谱数据库为参照, 结合课题组自主研发的 LIBSystem 软件自带的光谱数据库, 对 14 种钢铁的等离子体发射光谱进行采集与分析, 其光谱如图 2 所示。

建模前, 首先预选出 Mn, Ni, Cr, Mo 和 V 元素的共计 51 条待测谱线 (见表 2), 然后, 用 Fe: 404.58 nm 的谱线作为参考线, 对 51 条谱线进行归一化处理。预选谱线的原则是, 以 NIST 光谱数据库为依据, 同时挑选谱线强度较高、波形完整、没有自吸收现象和不被其他元素干扰的谱线作为分析线, 这样有利于建立多元素变量的分析模型来识别不同牌号的钢铁。

在统计学习中, 各类变量经常存在相关性导致输入向量的维度过高, 造成训练模型出现过拟合问题, 需要对输入变量进行降维处理。常见的降维方法有人工选取、PCA 和线性判别分析 (linear discriminant analysis, LDA) 等。但传统的降维方法常常会在降维的过程中丢失信息, 导致模型准确度不理想, 而人工选谱则过于麻烦, 测试结果有可能出现误判。测试覆盖度有限及人力成本有限是测试技术所面临的瓶颈。上述难题表明了对实现自动化选谱处理的渴望。本工作在分类前首先预选出可能的理想谱线, 然后通过计算机将预选谱线随机组合, 并且将这些组合作为输入特征建立多个 SVM 分类模型, 以寻找到一个理想可靠的模型对钢铁进行分类识别。

2.2 直接 SVM 分类结果

实验中, 采用交叉验证法获得 SVM 算法中惩罚因子 C

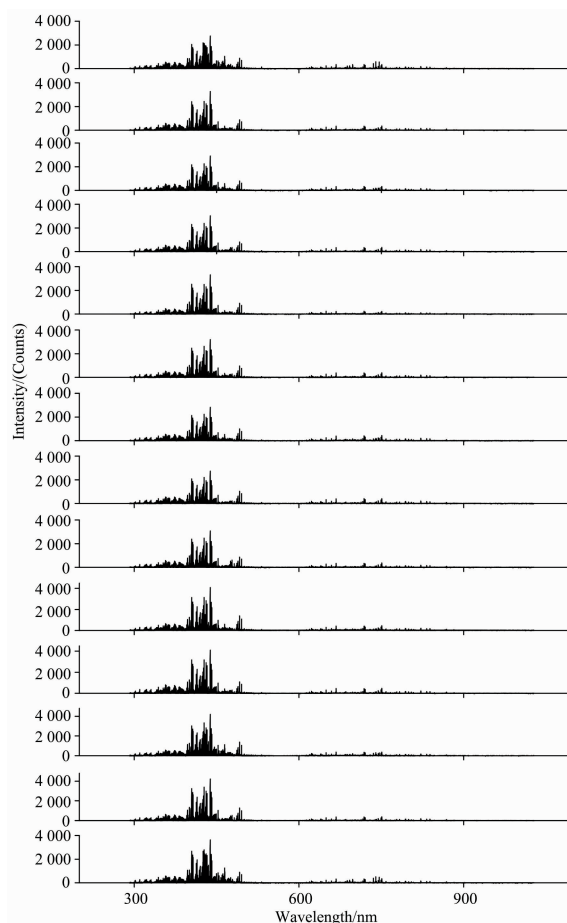


图 2 14 种特钢样品的等离子体发射光谱

Fig. 2 The emission spectra of 14 types of special steel samples

和核函数参数 g 的最优值, 由前文可知, 每个钢铁样品采集 30 组光谱数据, 随机选择其中 20 组作为训练集, 另 10 组作为测试集。因此训练集有 280 组, 测试集有 140 组光谱数据。

同时,依据表 1 的样品编号分别设置特钢的标签值为 1~14。首先,用训练集的 280 组光谱数据训练 SVM 模型,再将测试集中的 140 组光谱数据输入该 SVM 模型进行预测。预测结果如图 3 所示,图中,符号“ \circ ”代表每组光谱数据的实际标

签,符号“ \times ”代表预测的标签,当“ \circ ”与“ \times ”重合时,表示预测值与实际值一致;相反,“ \circ ”与“ \times ”不重合时,表示未能正确识别。

表 2 选择的特征谱线

Table 2 The selected emission lines

Element	Wavelength(λ /nm)
Ni	300.249, 301.200, 305.082, 310.156, 313.410, 341.476, 344.626, 345.846, 346.165, 349.296, 351.505, 352.454, 356.637, 361.939
Cr	357.868, 359.348, 425.433, 427.481, 428.973
Mn	380.672, 403.176, 403.307, 403.449, 404.136
Mo	315.817, 317.034, 319.398, 320.884, 344.712, 378.825, 386.410, 390.295, 406.988, 418.832, 441.169, 550.649, 553.303, 557.044
V	318.341, 318.399, 318.538, 370.357, 385.584, 390.226, 411.518, 412.806, 413.199, 437.923, 438.471, 439.522, 440.850

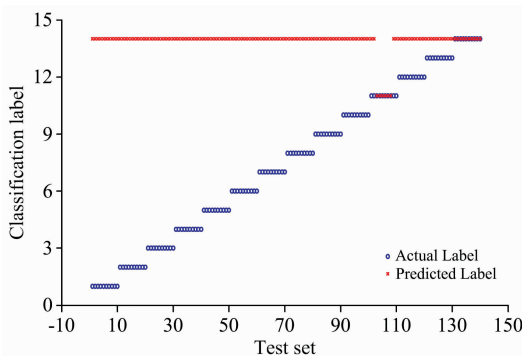


图 3 SVM 的预测结果

Fig. 3 The prediction results by SVM

由图 3 可见,当将 14 个钢铁样品的测试集 140 个光谱数据直接输入训练后的 SVM 模型时(选取的谱线为表 2 中的 Mn, Mo, V, Cr 和 Ni 的强谱线,共 51 条),图中出现了 124 个误判点,仅 11 号样品的分类效果较好,可能的原因是原始光谱的总体波动过大,导致预测结果趋于一个样品,即 14 号样品。而 11 号样品的波动对分类结果影响可能不是特别大,使得部分数据被正确预测。可见,采用 51 条金属元素的特征谱线作为 SVM 分类器模型的输入向量时,14 种钢铁的平均识别正确率仅为 11.43%,此时的模型并不理想。

2.3 归一化光谱数据后 SVM 分类结果

为了减小光谱数据波动的影响,进一步的提高 SVM 分类算法的准确率,以 Fe: 404.58 nm 的谱线作为参考线,对所选的待测元素的特征谱线光谱强度做归一化处理,将归一化后的光谱强度作为 SVM 分类模型的输入量进行训练和预测,预测结果如图 4 所示。结果表明,谱线强度做归一化处理后可以校正实验测量条件的扰动造成的偏差,减小实验条件波动的影响。

由图 4 可见,将归一化处理后的测试集共 140 个光谱数据输入训练后的 SVM 模型时,图中出现了 6 个误判点。由图可见,出现误判的点为 5, 6 和 11 号样品,根据表 1 各样品的元素浓度可知,5, 6 和 11 号样品所含 Mn, Cr, Ni, Mo

和 V 元素的浓度极其相似,例如,对 Cr 元素而言,6 号与 11 号 Cr 元素含量分别为 1.02% 和 1.05%;对 Ni 元素而言,5 和 6 号样品所含的浓度含量都为 0.02%;对 V 元素而言,5 和 6 号样品中所含的浓度含量分别为 0.006% 和 0.005%。以上元素浓度差别十分微弱,并且由于目前绝大多数的激光器能量并不十分稳定,导致某些样品中相似浓度的元素谱线做归一化后强度值极为接近,系统模型受此影响从而发生了误判。

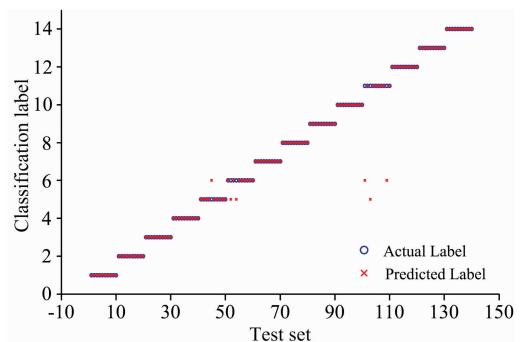


图 4 归一化后 SVM 的预测结果

Fig. 4 The prediction results of normalized spectra by SVM

采用归一化处理后的 51 条特征谱线作为 SVM 分类器模型的输入向量,14 种特钢的平均识别正确率为 95.71%,此时的分类模型虽能识别大多特种钢材,但精准度还并不理想。

2.4 谱线遍历组合后 SVM 分类结果

为了进一步提高分类的准确性,使用 MATLAB 程序遍历不同谱线的组合作为输入变量,进行多次建模,最终挑选出最优的输入特征,即 6 条最优特征谱线组合(如表 3 所示),此时的 SVM 判断准确率达到了 100%,其实验结果如图 5 所示。

表 4 为不同输入特征下的准确率与建模时间的比较。由表 4 可知,采用数据归一化处理+遍历组合选出的 6 条最优特征谱线作为输入向量时,SVM 模型识别准确度达到

表 3 SVM 预测准确度达 100% 的 6 条特征谱线
Table 3 The 6 spectral lines with SVM prediction accuracy of 100%

Element	Wavelength(λ /nm)
Mn	403.307
Mo	386.410
V	385.584
Cr	427.481, 357.868
Ni	352.454

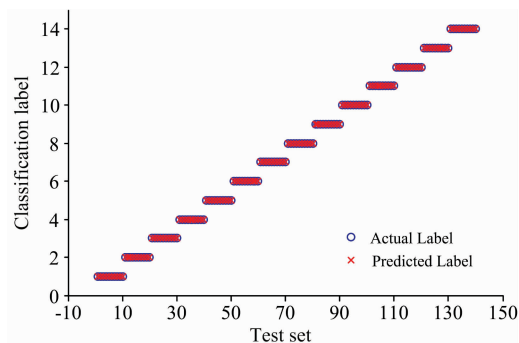


图 5 挑选 6 条特征谱线后 SVM 预测结果

Fig. 5 The SVM prediction results using selecting 6 special spectral lines

100%。这是由于在众多谱线中,算法通过遍历各种组合,只余下辨识度高、代表性强的谱线作为输入向量,使得模型的精准度得到大幅提升。从平均建模时间来看,对数据采用归一化+遍历组合的算法处理后,其建模时间也大大缩短,相比于使用原始光谱的 SVM 分类模型,其建模时间减小了 37%。

表 4 不同输入下 SVM 的平均预测准确率和平均建模时间
Table 4 The average prediction accuracy and mean modeling time of SVM with different inputs

Input	Average prediction accuracy/%	Mean modeling time/s
Preselected spectral data	11.43	0.0173 23
Normalized data	95.71	0.018 579
Optimal data of traversal	100	0.0108 56

3 结 论

基于光纤传能的移动式 LIBS 样机平台,采用预选谱线+遍历组合的降维方法与 SVM 算法相结合,对 14 个特钢材料的 LIBS 光谱进行快速分类。SVM 算法计算简单、训练速度快、并且每次训练的模型稳定。然而,当单独将 51 条预选谱线作为输入特征输入到 SVM 算法时,测试集中 140 组光谱数据在直接输入大量特征时发生了过拟合,导致模型的泛化程度降低,预测准确度仅为 11.43%,此时,对 14 种特钢材料的分类精度并不理想;当采用归一化光谱数据作为输入特征输入到 SVM 算法时,预测准确度提高到 95.71%,说明对光谱数据进行归一化处理可以显著提高 SVM 模型分类准确性;而当 51 条预选谱线经过谱线遍历组合降维之后,仅剩 6 条被选择的谱线作为输入特征时,此时的 SVM 模型预测准确度达到了 100%。实验结果表明,采用谱线遍历组合的方法是一种行之有效的降维方法。同时,该降维方法相对于人工选取特征谱线来说,方便快捷,操作简单,对模型的优化程度高。当面临大量特征数据时,机器自动选取特征与人工挑选谱线相比具有明显优势。可以看出,基于谱线遍历组合降维方法的 SVM 算法模型结合 LIBS 技术,在材料快速分类方面具有很好的工业应用前景。

References

- [1] Zeng Q D, Guo L B, Li X Y, et al. Journal of Analytical Atomic Spectrometry, 2016, 31(3): 767.
- [2] Sturm V, Fleige, Rüdiger, De Kanter M, et al. Analytical Chemistry, 2014, 86(19): 9687.
- [3] ZENG Qing-dong, ZHU Zhi-heng, DENG Fan, et al(曾庆栋,朱志恒,邓凡,等). Acta Photonica Sinica(光子学报), 2018, 47(8): 102.
- [4] Distjerdi M V, Mousavi S J, Soltanolkotabi, et al. Iran J. Sci. Technol. Trans. Sci., 2018, 42: 959.
- [5] Zeng Q D, Guo L B, Li X Y, et al. Journal of Analytical Atomic Spectrometry, 2015, 30(2): 403.
- [6] Zeng Q D, Deng F, Zhu Z H, et al. Plasma Science & Technology, 2019, 21(3): 034006.
- [7] Zou X H, Guo L B, Shen M, et al. Optics Express, 2014, 22(9): 10233.
- [8] Guo L B, Zhang B Y, He X N, et al. Optics Express, 2012, 20(2): 1436.
- [9] Hahn D W, Omenetto N. Applied Spectroscopy, 2012, 66(4): 347.
- [10] Yang Y W, Hao X J, Zhang L L, et al. Sensors, 2020, 20: 1393.
- [11] Kasem M A, Gonzalez J J, Russo R E, et al. Talanta, 2013, 108: 53.
- [12] YANG You-sheng, ZHANG Yan, YANG You-liang, et al(杨友盛,张岩,杨友良,等). Laser & Optoelectronics Progress(激光与光电子学进展), 2015, 52(5): 209.
- [13] GU Yan-hong, ZHAO Nan-jing, MA Ming-jun, et al(谷艳红,赵南京,马明俊,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(6): 1893.
- [14] Cortes C, Vapnik V. Machine Learning, 1995, 20(3): 273.
- [15] Zhang T, Wu S, Dong J, et al. Journal of Analytical Atomic Spectrometry, 2015, 30(2): 368.

Rapid Classification of Steel by a Mobile Laser-Induced Breakdown Spectroscopy Based on Optical Fiber Delivering Laser Energy

LI Wen-xin¹, CHEN Guang-hui^{1,3}, ZENG Qing-dong^{1,2*}, YUAN Meng-tian^{1,3}, HE Wu-guang¹, JIANG Ze-fang¹, LIU Yang¹, NIE Chang-jiang¹, YU Hua-qing¹, GUO Lian-bo²

1. School of Physics and Electronic-Information Engineering, Hubei Engineering University, Xiaogan 432000, China

2. Wuhan National Laboratory for Optoelectronics (WNLO), Huazhong University of Science and Technology, Wuhan 430074, China

3. Faculty of Physics & Electronic Sciences, Hubei University, Wuhan 430062, China

Abstract In order to realize the industrial on-site rapid detection and identification for special steel, a mobile laser-induced breakdown spectroscopy prototype based on optical fiber delivering laser energy is adopted in this experiment to collect the spectral data of 14 special sheets of steel. The spectra of special steels were rapidly classified via dimensionality reduction in which pre-selected spectral lines were traversed, combined with a support vector machine (SVM). In the experiment, original spectral data, normalized spectral data and normalized spectral data after traversed were used as the input vectors of the SVM classification model, and the recognition accuracy of the model for special steels under different input vectors was compared. The results show that on the basis that more than 51 spectral lines were selected as input variables, the recognition accuracy of normalized spectral data as input variables for steels reaches 95.71%. It is significantly higher than 11.43%, whose accuracy was used raw spectral data as the input vector. Further, the MATLAB program was used to traverse the spectral line combination to choose the optimal input features. When 6 specific spectral lines were selected, the accuracy of special steels recognition reached 100%, and the modeling speed was also improved accordingly. It can be seen that when a large number of common feature data are pre-selected, automatic feature selection by machine has obvious advantages over the spectral line of manual selection. The SVM algorithm based on this dimension reduction method has a good industrial application prospect in LIBS rapid classification technology.

Keywords Laser induced Breakdown spectroscopy; SVM; Spectral line traversal combination; Dimension reduction; Classification of steel

(Received Aug. 4, 2020; accepted Dec. 19, 2020)

* Corresponding author