

基于支持向量机的钢水 LIBS 定性分析

杨友盛¹ 张岩¹ 杨友良² 马翠红²

¹唐山赛福特智能控制股份有限公司研发中心, 河北 唐山 063000

²河北联合大学电气工程学院, 河北 唐山 063000

摘要 激光诱导击穿光谱(LIBS)技术具有快速、非接触、无需制样等特点, 适合应用于转炉钢水成分的在线分析。由于转炉终点可由 Si、Mn 含量和温度来判定, 因此提出了钢水成分中 Si 和 Mn 的 LIBS 定性分析方法。通过光谱仪采集激光激发的光谱, 经过预处理和寻峰等操作后, 以原子光谱数据库(NIST)为参考标准, 找出 Si 和 Mn 对应的特征谱线波长和光谱强度, 利用支持向量机(SVM)强大的分类功能和采集到的 245 组数据中的 210 组学习得到支持向量分类(SVC)模型, 利用 SVC 模型预测这 245 组数据, 结果证明该模型的准确率为 98% 以上, 将其应用在相同实验条件的情况下, 会大大减少 LIBS 定性分析时间。

关键词 光谱学; 激光诱导击穿光谱; 定性分析; 支持向量机; 特征谱线; 转炉终点

中图分类号 O433 文献标识码 A

doi: 10.3788/LOP52.053001

Qualitative Analysis of Molten Steel Based on SVM by LIBS

Yang Yousheng¹ Zhang Yan¹ Yang Youliang² Ma Cuihong²

¹Tangshan SOFT Intelligent Control Co., Ltd. R & D Center, Tangshan, Hebei 063000, China

²Hebei United University, College of Electrical Engineering, Hebei United University, Tangshan, Hebei 063000, China

Abstract Laser induced breakdown spectroscopy (LIBS) technology has the characteristics of speediness, non-contact, no need of sample preparation, which is very suitable for the online analysis of the converter steel composition, because the end-point can be determined by Si and Mn contents and temperature. A qualitative analysis of LIBS is proposed for analyzing the Si and Mn composition in molten steel. Laser excitation spectra are collected by spectrometer, and after the operations of pretreatment and peak searching, with the atomic spectra database (NIST) as the reference standard, the corresponding characteristics of spectral line wavelength and spectral intensity of Si and Mn are found out. Based on the powerful classification function of support vector machine (SVM), 210 sets of 245 sets of data collecting are used to get the support vector classification (SVC) model, which then predicts that 245 groups of data. The accuracy of the model is more than 98%, which can identify the corresponding wavelength of Si and Mn very well, and can be used under the condition of the same experimental conditions with significant reduction to the LIBS qualitative analysis time.

Key words spectroscopy; laser induced break-down spectroscopy; qualitative analytical; support vector machine; characteristic spectral line; end-point of basic oxygen furnace

OCIS codes 300.2140; 020.1335; 300.6210

1 引言

激光诱导击穿光谱(LIBS)^[1]是采用聚焦的高能量脉冲激光入射到样品的表面产生等离子体^[2], 通过分析等离子体从产生到湮灭过程中产生的原子辐射光谱, 从而推导出样品的元素组成成分及含量。LIBS 技术具有很多显著优点, 如能够探测所有元素, 可同时检测多种元素, 无需真空, 样品制备简单或无需制备, 被分析样品几乎无损, 可以实时分析、原位探测和远程探测等, 从而引起广泛注意, 并成为极具实用价值的光谱分析技术^[3]之一。近些年逐步向冶金行业发展。

收稿日期: 2014-10-29; 收到修改稿日期: 2014-11-20; 网络出版日期: 2015-04-23

基金项目: 国家自然科学基金项目(61271402)、国家自然科学基金(61171058)、国家科技部科技人员服务企业行动(2009GJA2000)

作者简介: 杨友盛(1975—), 男, 助理工程师, 主要从事冶金自动化研究方面的研究。E-mail: 602063812@qq.com

转炉炼钢是钢铁冶金中的一个重要环节,准确预测转炉炼钢终点^[4]决定了出炉钢铁的质量。传统的预测方法是离线检测钢水成分,不仅耗时长,而且还增加了炼钢成本,不能做到实时的在线监测,是目前钢铁行业亟需解决的难题。由于转炉钢铁中 Si、Mn 含量和温度存在着一定的变化规律,因此,能够准确测量 Si 和 Mn 的含量就可以预测转炉终点,由于 LIBS 技术水平还不够成熟,在工业应用中还比较少,特别是钢铁行业环境复杂,很多研究工作还停留在实验室阶段。谢承利^[5]将 LIBS 技术应用在煤的成分检测中取得了不错的效果,大大减少了传统煤成分检测时间;张旭等^[6]将其应用于检测蔬菜水果的重金属含量是否超标,但只限于固定几种重金属元素的半定量分析;卢伟业等^[7]利用 LIBS 技术检测复合肥重氮磷钾含量,也只是对几种特定的元素进行检测,并且以上都是在常温下检测,没有外界环境条件的影响。

支持向量机^[8]是近年来兴起的一种机器学习方法,包括支持向量回归(SVR)和 SVC,它可以解决小样本、非线性、高维数、局部极小点等实际问题,已经在模式识别、回归预测等领域取得了较成功的应用,尤其是在解决小样本和非线性问题方面有其他方法无法比拟的优势。本文用到了支持向量机的分类功能,对光谱数据进行定性分析。由于支持向量机也有很好的回归预测功能,因此课题在后续的工作中可以将支持向量回归用于元素的定量分析中,提高定量分析精度,做到实时在线监测炉内情况,为炼钢过程作出必要的指导,提高炼钢命中率。将支持向量分类运用于转炉终点预测中,将 Si、Mn 光谱强度和温度定为 3 个分类属性用于预测是否达到终点。

文章采用支持向量机的分类功能^[9],利用波长和对应的光谱强度作为两个分类属性变量,将变量分为 3 类:Si、Mn 和其他元素。利用采集的数据学习得到分类模型,利用分类模型进行定性分析。

2 实 验

实验装置的结构图如图 1 所示。等离子激发源为德国 INNOLAS 的 Nd:YAG 激光器,其输出波长为 1064 nm,最大输出能量为 400 mJ,脉冲重复率为 10 Hz,脉冲宽度 5 ns,脉冲激光器通过焦距为 50 cm 的凸透镜汇聚在样品表面产生等离子,光纤探头以一定的角度固定在一个旋转电位移台上,当实验开始时通过位移台旋转至炉口上方 10 cm。采用海洋光学公司的 HR2000+型光谱仪采集激发后的光谱图像,运用自制的时序控制器^[10]来控制激光器与光谱仪采集之间的时序关系,光谱仪分辨率为 0.1 nm,测量波长范围为 200~630 nm。实验材料采用某钢厂的 132E 钢,其主要成分如表 1 所示。

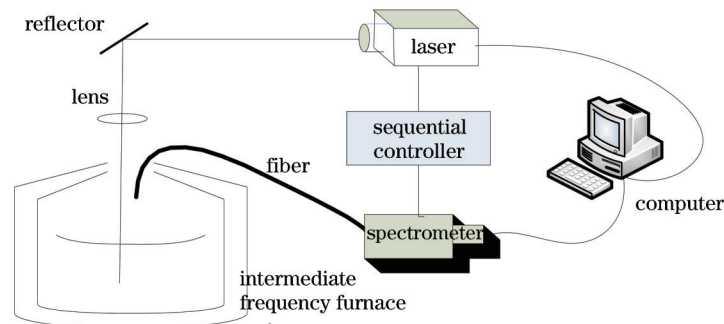


图 1 实验装置结构图

Fig.1 Structure diagram of experimental apparatus

表 1 132E 钢的主要成分含量

Table 1 Main components content of 132E steel

Element	C	Si	Mn	P	S	Cu	Cr	Ni
Mass fraction /%	6.4	4.8~3.2	3.9	1.2	0.6	0.9	0.7	0.6

如图 1 所示,利用中频炉熔化钢块,在激光器发射激光前,必须采集多组背景光的光谱,以便在得到的激发光谱中扣除背景光。

3 支持向量机分类模型

SVM 理论是从线性可分的最优分类面发展而来的,而对于现实的非线性问题,是将输入向量经过某一非

线性映射(该非线性映射函数就是核函数),将其映射到高维特征空间中。在这一特征空间中,构造一个最优分类超平面^[11],该超平面不仅能将两类正确分开,而且要使分类间隔最大,如图2所示是一个两类样本的情况。

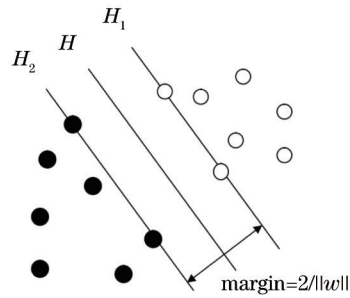


图2 最优分类超平面

Fig.2 Optimal hyper plane

图2中,通过核函数映射的样本分为两类, H 为最优分类超平面,它距离两类样本的间隔最大,而 H_1 和 H_2 上的4个样本点就是支持向量, H_1 表示 $wx + b = 1$, H_2 表示 $wx + b = -1$, H 表示 $wx + b = 0$,最大分类间隔为 $2/\|w\|$ 。由于支持向量机分类模型是将样本数据分为两类,因此本文中的分3类的情况就需要两次分类计算。第一次分类算法将Si和Mn分为一类,其他的元素分为一类,设这两类样本集为 (x_i, y_i) , $x_i \in R^n$, $y_i \in \{-1, 1\}$, $i = 1, 2, \dots, N$ 。其中 N 为训练样本总数, n 为样本空间的维数, y 为样本的类别标志,根据结构风险原则,分类问题的最小风险界为

$$\min R(w, \xi) = \frac{1}{2} w^T w + c \sum_{i=1}^N \xi_i \quad (1)$$

式中 w 为超平面的斜率, ξ 为样本点错分程度, c 为惩罚因子。

经过拉格朗日函数优化,将其转化成在条件 $0 \leq \alpha_i \leq c$ (c 为常数)和 $\sum_{i=1}^n \alpha_i y_i = 0$ 约束下,对拉格朗日乘子 α_i 求解函数的最大值

$$Q[\alpha, \phi(x_i)] = -\frac{1}{2} \sum_{i,j=1}^N y_i y_j \phi(x_i)^T \phi(x_j) \alpha_i \alpha_j + \sum_{i=1}^N \alpha_i \quad (2)$$

此为在不等式约束下的一个二次规划问题,有唯一解。核函数 $K(x_i, x_j)$ 满足Mercer条件:

$$\varphi(x_i)^T \varphi(x_j) = K(x_i, x_j) \quad (3)$$

所以, α_i 即为下列二次规划问题的解

$$\max Q_1[\alpha, K(x_i, x_j)] = -\frac{1}{2} \sum_{i,j=1}^N y_i y_j K(x_i, x_j) \alpha_i \alpha_j + \sum_{i=1}^N \alpha_i \quad (4)$$

式中只有少部分 α_i 不为零,与之对应的样本就是支持向量。因此可以得到最后的最优分类决策函数

$$f(x) = \text{sign} \left[\sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b^* \right] \quad (5)$$

式中 b^* 分别为超平面的分类阈值。经过上述步骤,已经将样本分成两类,分为Si、Mn和其他元素两类,接下来再经过一次上述的分类办法,就可以将Si和Mn分开。

4 光谱数据处理与分析

4.1 特征谱线分析

图3是从实验得到的LIBS光谱数据中抽出的一个经过预处理的光谱数据,分析可知,钢水的LIBS光谱数据非常复杂,由于钢水中含量最多的是Fe元素,元素的特征谱线都会受到基体效应^[13]的影响,绝大部分元素的特征谱线都会被铁基体的谱线淹没。实验中得到的特征谱线一般都是元素的灵敏线和次灵敏线,经过谱线自动寻峰,可以得到特征谱线及其对应的相对强度。利用美国国家标准与技术研究院(NIST)的原子发射光谱数据库,结合实验得到的数据,通过MATLAB自寻峰,经过对比峰值对应的波长,就可以得到样品中元素对应的特征谱峰。表2为样品中元素在NIST原子发射光谱数据库中对应的特征波长,样品中C、P等非

金属元素激发困难,其特征谱线被基体效应淹没。实验选择了精度比较高的光谱仪,精度达到 0.1 nm,有效地限制了谱线重叠问题,光谱仪本身暗电流较小,谱线漂移情况小,并且在比较特征谱线时,在实验得到的峰值波长前后 0.25 nm 范围内来对应 NIST 数据中特征谱线波长,当有多个 NIST 特征谱线在这个范围内时,结合钢中所含元素,取距离小的特征谱线。

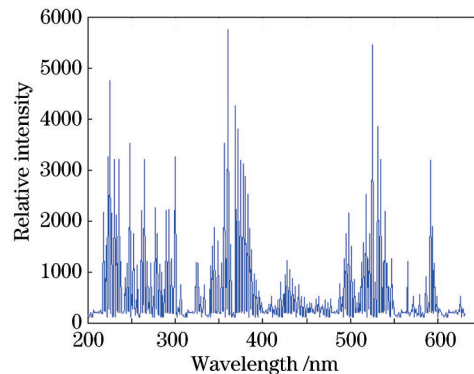


图3 LIBS实验数据

Fig.3 Experimental data of LIBS

文章采取的自动寻峰^[14]方法是二阶微商的办法,通过对谱线进行二次求倒,得到的图像在 y 轴负方向上的极小值就是谱峰。对于原谱线基底上的弱峰,因其变化较为缓慢,二阶微商后会变成幅值很小的峰或者平滑线,如果设定一个合适的阈值(靠经验来设定,论文中的阈值取光谱最高强度的 1/6),那么就能很清楚地分辨出谱线峰值。

表2 样品元素中检测到的特征谱线

Table 2 Theory of characteristic line of sample element

Element	Characteristic wavelength /nm
C	Undetectable
Si	250.69 251.43 251.61 252.41 252.85 288.16 390.55
Mn	257.61 259.37 260.57 279.48 279.83 280.11 403.08 403.31 403.45
P	Undetectable
S	Undetectable
Cu	219.23 224.70 324.75 327.40 510.55 515.32 521.82
Cr	283.56 284.33 284.98 285.57 286.09 357.87 359.35 360.53 425.43 427.48 428.97 520.45 520.60 520.84
Ni	225.39 226.45 227.02 228.71 300.25 305.08 341.48 346.17 349.30 351.51 352.45 361.94

4.2 实验数据分析

利用支持向量机分类功能来建立数据分类模型,而由于支持向量机分类实质上还是属于一种统计学方法,因此,需要制备数据样本。实验得到了多组实验数据,提取了其中的 245 组,其中 80 条 Si 的特征谱线,80 条 Mn 的特征谱线,其他元素一共 85 条,这些样本数据尽量包含所有能在实验中检测出来的 Si 和 Mn 元素特征谱线,这样有利于建立更加精确的分类模型。在标签矩阵中, Si 定义为 1, Mn 定义为 2,其他的元素定义为 3。先利用 245 组样本数据中的 210 组来建立预测分类模型,然后用所有的 245 组数据来验证分类模型的准确度。

支持向量机分类的基本思想是经过一个映射关系,将样本数据映射到一个高维空间,在这个高维空间中存在一个超平面可以将不同属性的样本分开,而这个映射关系就是核函数,因此必须选择合适的核函数,一般情况下,常用的分类核函数有 3 种,由于径向基神经网络核函数表示形式简单,有多个变量输入时不会增加太多复杂性,且径向对称、任意阶导数存在,有利于进行理论分析,因而选择其中径向基神经网络核函数^[13] $K(x_i, x) = \exp(-\sigma |x - x_i|^2)$ 作为分类器核函数,经过推导可得相应的决策函数为

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right], \quad (6)$$

式中 α_i^*, b^* 为上面二次规划问题的解。

建立分类模型的基本步骤如下:

1) 确定训练样本数据, 实验中为 210 组, 其中 Si、Mn 以及其他元素特征谱线各选 70 组, 建立测试数据矩阵和标签矩阵, 并对测试数据进行归一化处理。

2) 在 MATLAB 下安装 SVM 工具箱, 在数据处理过程中可调用工具箱函数, 选择神经网络核函数, 以及合适的惩罚参数 $c > 0$, 在选择合适的惩罚参数时, 可选择一个范围, 按一定的步距进行求解, 选择训练准确率最高并且惩罚参数值最小的数。

3) 利用 MATLAB 编写函数解决最优分类超面对应的二次规划问题, 并求解。

4) 构造最优分类超平面, 并得到决策函数, 将测试样本代入就可得到生成的测试标签矩阵和测试的分类准确率。

经过训练的模型必须达到参数最优的效果, 实验中采用网格参数寻优(将两个参数按一定的步长组成多组数据来计算最优解)方法, 通过交叉检验的方法(利用允许的不同参数建立模型, 此模型将所有的样本数据预测一次, 求所有误差的平方和, 当平方和最小时, 说明这一组参数是最优的), 得到分类准确率最高的模型, 有时候多组参数对应的分类准确率都是最高的, 这时就要选择惩罚函数最小的那组。图 4 为参数寻优结果的等高线图, 显示的是不同 c 和 g 的组合得到的分类准确率, 图中最高的分类准确率达到 98.8%, 由于选择的核函数是径向基神经网络核函数, 因此在 MATLAB 支持向量机工具箱中 g 被定义为 1, 可得最佳的惩罚函数 c 为 1.414。

确定好参数后, 利用所有的 245 组数据对模型进行准确性检验, 得到如图 5 所示的实际分类与预测分类比较图。

由图 5 可以看出, 实际分类与预测分类只有 3 处不一样, 准确率达到 98.8%, 完全可以用来进行光谱定性分析, 分析这 3 处错误的原因可能是由于这 3 条特征谱线强度弱, 并且在一组实验数据中大部分的特征谱线是属于 Si 和 Mn 以外的特征谱线, 总的其他元素的样本特征谱线为 85 条, 可能覆盖得不是很完全, 学习得到的模型预测时就会不准确, 而且这 3 条谱线在 250~280 之间, 而在这一波段 Si 和 Mn 的特征谱线分布比较密集, 因此会给模型带来一定的误差。

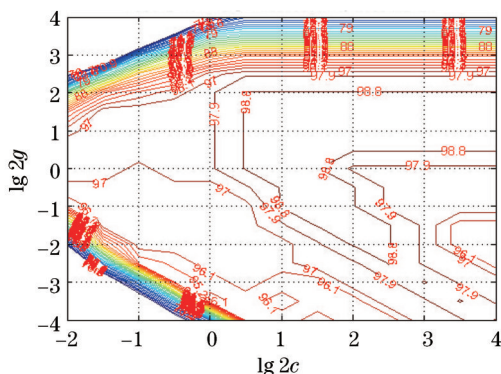


图 4 SVC 参数选择结果图

Fig.4 Parameter selection results of SVC

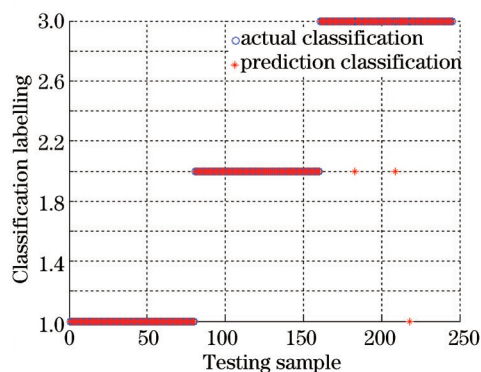


图 5 实际分类与预测分类图

Fig.5 Comparison between actual classification and prediction classification

5 结 论

利用支持向量机得到的分类模型能够很好地分辨出 LIBS 实验中得到的特征谱线是属于哪一元素的特征谱线, 准确率能够达到 98.8% 以上, 在实验环境不变的情况下, 可以用在光谱实验定性分析上。通过实验, 得到支持向量机对于光谱实验这种小样本数据有很好的学习功能, 而且支持向量机还有回归预测的功能, 如果能构造合适的核函数, 完全可以用在定量分析上, 这为今后在定量分析的问题上指明了一个方向。

参 考 文 献

- 1 C Pasquini, J Cortez, L M C Silva, *et al.*. Laser induced breakdown spectroscopy[J]. Journal of the Brazilian Chemical Society, 2007, 18(3): 463-512.

- 2 Meng Fanli. Research on Automatic Selection of Analytical Lines in Laser- Induced Breakdown Spectroscopy[D]. Tianjin: Tianjin University, 2012: 3-5.
孟繁莉. 激光诱导击穿光谱的分析线自动选择方法研究[D]. 天津: 天津大学, 2011: 3-5.
- 3 Zhang Wenyan, Lin Zhaoxiang, Lu Jidong, *et al.*. Investigation on laser induced breakdown spectroscopy of potassium chloride solution[J]. Applied Laser, 2007, 27(2): 121-123.
张文艳, 林兆祥, 陆继东, 等. 氯化钾溶液的激光诱导击穿光谱研究[J]. 应用激光, 2007, 27(2): 121-123.
- 4 Yang Youliang, Zhan Yan, Meng Fanwei. Prediction model of converter steelmaking endpoint based on spectral in formation and temperature[J]. Metallurgical Analysis, 2012, 32(11): 1-5.
杨友良, 张 岩, 孟凡伟. 基于光谱信息与温度的转炉炼钢终点预测模型[J]. 冶金分析, 2012, 32(11): 1-5.
- 5 Xie Chengli. Study of the Spectral Data Processing in Laser Induced Breakdown Spectroscopy Analysis and Its Application in Elemental Analysis of Coal[D]. Wuhan: Huazhong University of Science and Technology, 2009.
谢承利. 激光诱导击穿光谱数据处理方法及在煤分析中的应用研究[D]. 武汉: 华中科技大学, 2009.
- 6 Zhang Xu, Yao Mingyin, Liu Muhua, *et al.*. Quantitative analysis of chromium in kelp by laser- induced breakdown spectroscopy[J]. Acta Agriculturae Universitatis Jiangxiensis, 2012, 34(1): 187-190.
张 旭, 姚明印, 刘木华, 等. 海带中铬含量的激光诱导击穿光谱研究分析[J]. 江西农业大学学报, 2012, 34(1): 187-190.
- 7 Lu Weiye, Lu Jidong, Yao Shunchun, *et al.*. Synchronozation detection of nitrogen phosphorus potassium in compound fertilizer with laser induced breakdown spectroscopy[J]. Chinese J Lasers, 2011, 38(10): 1008003.
卢伟业, 陆继东, 姚顺春, 等. 复合肥氮磷钾元素含量的激光诱导击穿光谱同步测量[J]. 中国激光, 2011, 38(10): 1008003.
- 8 Li Xuehua, Shu Lan. Fuzzy theory based on support vector machine classifier[C]. The 5th International Conference on Fuzzy System and Knowledge Discovery, 2008: 600-604.
- 9 Kwok T J. Support vector mixture for classification and regression problems[C]. The 14th International Conference on Pattern Recognition, 1998: 255-258.
- 10 Yuan Liangjing, Shi Xiaoxi, Lian Zhiqiang, *et al.*. Study on the improvement of analysis precision for laser induced breakdown spectrometry[J]. Metallurgical Analysis, 2012, 32(2): 1-5.
袁良经, 石小溪, 连志强, 等. 激光诱导击穿光谱法分析精度提高的研究[J]. 冶金分析, 2012, 32(2): 1-5.
- 11 S Abe. Analysis of multi-class support vector machines[C]. International Conference on Computational Intelligence for Modeling, Control and Automation, 2003: 385-396.
- 12 Wang Dongli. Scalable SVM Classifier with Application in Credit Rating[D]. Shanghai: Donghua University, 2010.
王冬丽. 基于可扩展的支持向量机分类算法及在信用评级中的应用[D]. 上海: 东华大学, 2010.
- 13 J A Aguilera, C Aragon, V Madurga, *et al.*. Study of matrix effects in laser induced breakdown spectroscopy on metallic samples using plasma characterization by emission spectroscopy[J]. Spectrochimica Acta Part B, 2009, 64(10): 993-998.
- 14 Wu Shaobo, Ye Lianhui. Automatic recognition method of element emission lines in composiiton analysis with LIBS [J]. Chinese J Scientific Instrument, 2014, 35(3): 670-675.
吴少波, 叶连慧. LIBS成分分析中元素谱线的自动识别方法[J]. 仪器仪表学报, 2014, 35(3): 670-675.
- 15 Zhang Ling. The theory of SUM and programming based learning algorithms in neural networks[J]. Chinese J Computers, 2001, 24(2): 113-118.
张 玲. 支持向量机理论与基于规划的神经网络学习算法[J]. 计算机学报, 2001, 24(2): 113-118.

栏目编辑: 张 雁