Spectrochimica Acta Part B 107 (2015) 45–55

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

journal homepage: www.elsevier.com/locate/sab

Laser-induced Breakdown spectroscopy quantitative analysis method via adaptive analytical line selection and relevance vector machine regression model

Jianhong Yang [a,*], Cancan Yi [a], Jinwu Xu [a], Xianghong Ma [b]

[a] School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, 100083, China
[b] School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK

ARTICLE INFO

Article history:
Received 17 August 2014
Accepted 21 February 2015
Available online 3 March 2015

Keywords:
Laser-induced breakdown spectroscopy
Quantitative analysis
Analytical line adaptive selection
Relevance vector machine regression model

ABSTRACT

A new LIBS quantitative analysis method based on analytical line adaptive selection and Relevance Vector Machine (RVM) regression model is proposed. First, a scheme of adaptively selecting analytical line is put forward in order to overcome the drawback of high dependency on a priori knowledge. The candidate analytical lines are automatically selected based on the built-in characteristics of spectral lines, such as spectral intensity, wavelength and width at half height. The analytical lines which will be used as input variables of regression model are determined adaptively according to the samples for both training and testing. Second, an LIBS quantitative analysis method based on RVM is presented. The intensities of analytical lines and the elemental concentrations of certified standard samples are used to train the RVM regression model. The predicted elemental concentration analysis results will be given with a form of confidence interval of probabilistic distribution, which is helpful for evaluating the uncertainness contained in the measured spectra. Chromium concentration analysis experiments of 23 certified standard high-alloy steel samples have been carried out. The multiple correlation coefficient of the prediction was up to 98.85%, and the average relative error of the prediction was 4.01%. The experiment results showed that the proposed LIBS quantitative analysis method achieved better prediction accuracy and better modeling robustness compared with the methods based on partial least squares regression, artificial neural network and standard support vector machine.

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a type of atomic emission spectroscopy that has well-known advantages of low requirement of sample preparation, micro-destructiveness, short measurement time and suitability of multi-element analysis in situ [1–3]. In the research of the qualitative and quantitative analysis of LIBS [4–6], the analysis accuracy has always been an important issue. The efforts to improve the quantitative analysis accuracy can be divided into two ways. One way is to calibrate the intensity of the LIBS signal obtained in the measurement. For example, Wang et al. [7] proposed a practical spectrum standardization approach to alleviate the negative effects caused by uncertainties of LIBS signals. However, nonlinear relationships between spectral line intensity and plasma temperature and electron number concentration make it still a challenge to achieve high accuracy even if the given spectral line intensities are absolutely correct. As a result, the other way, that is to establish more robust and precise mathematical model to describe the relationship between the spectral line intensity and the elemental concentration, has attracted more and more attention.

The LIBS quantitative analysis methods have two categories: the method with calibration curves of standard samples and the calibration-free method. In the calibration-free quantitative analysis method, the elemental concentrations are calculated by the spectral line intensities and experimental parameters based on the Boltzmann model. However, because of the hypotheses of local thermodynamic equilibrium (LTE) in the temporal and spatial observation window, the spatial homogeneousness of the radiation source and the optically thin plasma, the calibration-free (CF) quantitative analysis accuracy greatly depends on the accuracy of the spectral line intensity and model parameters and is significantly affected by self-absorption effect. There is still a lot of work needed to develop the CF-LIBS methodologies [8,9]. The main idea of quantitative analysis method with calibration curves of standard samples is to establish a statistical regression model between the elemental concentrations and spectral intensities. The parameters of the regression model will be trained by certified standard samples. Once the statistical regression model is obtained, the concentration of the analyte will be given by the regression model after taking the spectral intensities of analytical lines as the model input variables. The method with calibration curves of standard samples has been widely applied in the area of coal industry [10], mineral analysis [11], metallic elements detection [12] etc. However, there are still two

* Corresponding author. Tel.:+86 10 62332329; fax: +86 10 62332329.
E-mail address: yangjianhong@me.ustb.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.sab.2015.02.014
0584-8547/

challenges in achieving high analysis accuracy and model robustness. One is the selection of the analytical lines from hundreds of characteristic spectral lines; the other is the model's nonlinear generalization ability for limited training samples with uncertainties of spectral intensities.

Due to the large number of emission lines, choosing the proper lines as the input variables of the quantitative regression model has a great influence on the accuracy of the analysis results. The more analytical lines are chosen, the more complex the regression model will be. However, insufficient analytical lines will lead to poor analysis accuracy, because the information contained in the analytical lines are not adequate for representing the statistical characteristics of the atomic emission during the LIBS measurement. Most of the current analytical lines selection methods mainly rely on empirical principles. Mohamed [13] proposed that there were three rules which should be used for selecting the appropriate wavelength for each element to avoid the problems of self-absorption and spectral line interference. First, interference between spectral lines for different species must be avoided when selecting the analytical lines. Second, the possibility of self-absorption in the case of lines ending in a heavily populated level should be avoided, such as resonant lines. Third, the selected analytical lines should have similar excitation level in order to reduce the influence of the matrix effect. Laville et al. [14] preferred to select lines with low overlapping interference and low self-absorption effect by experimental trials. However, the above methods relied on prior knowledge and the selection criteria are usually subjective and vague so that automatic analytical lines selection cannot be implemented. Thus, it is always a heavy workload for selecting proper analytical lines from numerous spectral lines. Moreover, since there is no uniform selection rules, the analysis accuracy can be improved only be trial and error. In 2012, Du [15] proposed a method for automatically selecting the analytical lines based on two parameters: the relative detected-to-reference intensity ratio and the wavelength difference of the detected and theoretical lines. However, the reference intensity value which was used to define the relative detected-to-reference intensity ratio was simply determined by the relative intensity found in the NIST database [16], ignoring the differences between the practical experimental conditions and the conditions considered in NIST database. Additionally, in Du's method the procedure of searching spectral line interference in the range adjacent to the candidate analytical line was also dependent on prior knowledge. Nevertheless, since the experimental conditions of standard samples for model training and the samples to be analyzed may be different, it means that for different samples the analytical lines selected from the same procedures and criteria may not be identical. A method that can adapt to not only the variety of characteristics of spectral lines but also different samples under different measurement conditions is required for the selection analytical lines in practical applications. In this paper, a new analytical line adaptive selection method is proposed in which the built-in spectral characteristics (such as intensity, wavelength and the spectral peak width) are used to automatically select spectral lines by threshold screening procedures, and the intersection of the selected candidate spectral lines of all samples for both model training and testing will be determined as the analytical lines.

After the analytical lines are determined, building a statistical regression model between the spectral intensities and the elemental concentration is the critical step to ensure the accuracy of LIBS quantitative analysis. The main difficulties come from the following: (1) the relationship between the spectral intensities and the elemental concentration is complex and nonlinear due to the self-absorption effect, matrix effects, etc.; (2) in practical applications, high requirement for the regression model's generalization ability always conflicts with the fact that only a small amount of standard samples can be used in model training; (3) LIBS has a characteristic of poor reproducibility, as a result it needs a highly robust regression model for quantitative analysis; (4) the measured spectral intensity could be affected by the disturbance of the stray light from the background, the uncertainty of the spectral intensities during the process of shot-to-shot measurements and the noise of the spectrometer instrument. For a regression model, if the input contains noise, the output will contain noise as well. Thus, when the spectral intensities with noise are input into the regression model, as the output or the prediction result of the model, the elemental concentration will contain a portion of uncertainty which can be described by a certain probability distribution. This kind of uncertainty is related to the spectral intensity's noise among training samples and the functional fitting ability of the regression model. Quantitatively evaluating the uncertainty of a regression model's output is helpful when making a decision on to what extent the analysis result should be trusted. Therefore, an excellent regression model for LIBS quantitative analysis should consider not only the nonlinear fitting ability and the generalization ability for limited samples, but also the model robustness for noisy input and the ability of evaluating the uncertainty of the predicted output. The commonly used methods for the regression model of LIBS quantitative analysis include the partial least squares (PLS) [17,18], the artificial neural network (ANN) [19,20] and the standard support vector machine (SVM) [21–23]. The PLS model has shortcomings when representing the nonlinear relationship between the spectral intensities and the elemental concentration because in essence it is a linear regression method. ANN has the ability of nonlinear modeling, but it requires a large number of training samples and is easy to fall into local optima during error training, also the model prediction ability is limited when only a small number of the standard samples can be used. The SVM model is a type of structural risk minimization model which is suitable for regression modeling of limited samples, but it cannot give the probability distribution of the model prediction result, also the support vectors in the model are not sparse which means the generalization ability is limited. Relevance vector machine was firstly proposed by Tipping [24] in 2001 and has been successfully applied in the field of machine learning as a sparse probabilistic model based on limited samples [25,26]. RVM not only has excellent fitting and generalization ability but also can give a probability distribution of the prediction result and is more robust for noisy samples. Hence, the RVM model to the LIBS quantitative analysis is introduced here to improve the analysis accuracy and modeling robustness.

This work mainly focuses on the quantitative analysis method of calibration-based LIBS. A method of analytical lines adaptive selection is proposed so as to reduce the dependence of empirical principles and prior knowledge. Meanwhile, an LIBS quantitative regression model based on RVM is proposed, which not only retains the nonlinear generalization ability as the standard SVM for limited samples but also provides the ability of evaluating the model's prediction error. Experimental test of 23 alloy samples has been conducted and the results showed that the average relative error of the predicted Cr concentration was up to 4.01% which fully satisfied the requirements of industrial application.

The rest of the paper is organized as follows. In Section 2, the basic ideas of the adaptively selecting analytical lines and quantitative analysis model based on RVM are introduced. The instrumentation and experimental scheme are described in Section 3. Calculation and analysis results are listed in Section 4, and discussions are included in Section 5. The final conclusions are given in Section 6.

2. Theory

Within a certain temporal and spatial range, provided that the plasma is optically thin and in local thermal equilibrium and the effect of self-absorption can be neglected, the measured LIBS spectral line intensity can be expressed as [27]

$$I_{\lambda}^{kj} = FC_s \frac{A_{kj} g_k}{U_s(T)} e^{-E_k/(k_B T)} \tag{1}$$

where λ is the wavelength corresponding to the transition between two energy levels, $U_s(T)$ is the partition function of the emitting species s, E_k