



Technical Note

Comparative investigation of partial least squares discriminant analysis and support vector machines for geological cuttings identification using laser-induced breakdown spectroscopy[☆]

Ye Tian^a, Zhennan Wang^a, Xiaoshuang Han^{a,b}, Huaming Hou^a, Ronger Zheng^{a,*}^a Optics and Optoelectronics Laboratory, Ocean University of China, Qingdao, Shandong 266100, China^b College of Electronic Information Engineering, Inner Mongolia University, Hohhot, Inner Mongolia 010021, China

ARTICLE INFO

Article history:

Received 30 December 2013

Accepted 20 October 2014

Available online 30 October 2014

Keywords:

Laser-induced breakdown spectroscopy (LIBS)

Cuttings identification

Partial least squares discriminant analysis

(PLS-DA)

Support vector machines (SVM)

ABSTRACT

With the hope of applying laser-induced breakdown spectroscopy (LIBS) to the geological logging field, a series of cutting samples were classified using LIBS coupled with chemometric methods. In this paper, we focused on a comparative investigation of the linear PLS-DA method and non-linear SVM method. Both the optimal PLS-DA model and SVM model were built by the leave-one-out cross-validation (LOOCV) approach with the calibration LIBS spectra, and then tested by validation spectra. We show that the performance of SVM is significantly better than PLS-DA because of its ability to address the non-linear relationships in LIBS spectra, with a correct classification rate of 91.67% instead of 68.34%, and an unclassification rate of 3.33% instead of 28.33%. To further improve the classification accuracy, we then designed a new classification approach by the joint analysis of PLS-DA and SVM models. With this method, 95% of the validation spectra are correctly classified and no unclassified spectra are observed. This work demonstrated that the coupling of LIBS with the non-linear SVM method has great potential to be used for on-line classification of geological cutting samples, and the combination of PLS-DA and SVM enables the cuttings identification with an excellent performance.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In the geological logging field, the identification of cuttings plays an important role in the interpretation of an underground terrane and its content of oil, gas and water [1,2]. To improve the efficiency of the logging procedure, several new cutting identification methods have been developed based on digital image processing [3], spectroscopic analysis [4], and more recently, laser-induced breakdown spectroscopy (LIBS) [5,6]. For LIBS-based classification, in order to fully and rapidly utilize the information from the broadband LIBS spectra, the uses of multivariate analysis in the classification procedure have increased in recent years [7–12]. However, most of these studies rely on classical chemometric methods that are constrained by the underlying linearity treatment, such as principal component analysis (PCA) [8–10], soft independent modeling of class analogy (SIMCA) [9,10], and partial least squares discriminant analysis (PLS-DA) [8–12]. These linear classification methods may not be suitable to address the non-linear relationships in LIBS spectra, which could be caused by self-absorption in the plasma, matrix effect, and unstable ablation of the sample [13].

Sirven et al. found that the artificial neural network (ANN) method achieved better results than PLS since ANN could deal with non-linear relationship [14]. Wang et al. proposed a non-linearized PLS model which showed a significant improvement compared with the conventional PLS model [15].

Support vector machines (SVM) is a relatively new non-linear chemometric method that has been preliminarily applied in LIBS measurements, including the quality control of pharmaceutical process [13], and the classification of pigments and inks [16]. In this paper we focus on a comparative investigation of the conventional PLS-DA method and non-linear SVM method, for the LIBS-based identification of geological cutting samples. We also design a new classification approach by the joint analysis of PLS-DA and SVM, with the hope of further improving the classification accuracy.

2. Experiment

The schematic of the LIBS experimental setup is shown in Fig. 1. The details of the LIBS system were described previously [5], and are briefly described here. The plasma generation was carried out using a Q-switched Nd:YAG laser (Quantel Brilliant B) operating at 1064 nm. The laser pulse duration was 10 ns. We used low laser energy of 15 mJ to avoid cutting shakiness caused by the laser ablation. A part (5%) of each laser pulse passing through the beam shutter (BS in Fig. 1) was

[☆] Selected Paper from the 7th Euro-Mediterranean Symposium on Laser Induced Breakdown Spectroscopy (EMSLIBS 2013), Bari, Italy, 16–20 September 2013.

* Corresponding author. Tel.: +86 532 66781211.

E-mail address: rzheng@ouc.edu.cn (R. Zheng).

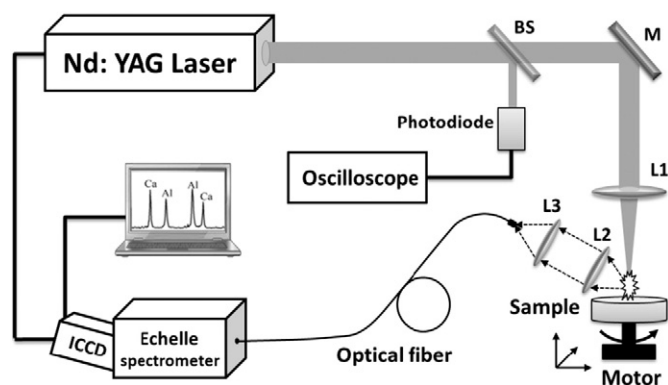


Fig. 1. Schematic of the experimental setup. BS: beam splitter; M: mirror; L1, L2, and L3: lenses.

sent to a photodiode connected with an oscilloscope, which monitored the fluctuation of laser energy. The laser beam was focused onto the sample with a UV fused silica plano-convex lens (L1, focal length 100 mm). The sample was rotated by a step-motor with a revolving speed of 30 r/min, and in this case we can sample a fresh spot randomly after each laser shot. The emission from the plasma was collected by a pair of identical UV fused silica plano-convex lenses (L2 and L3, focal length: 38.1 mm). The angle between the normal to the sample stage surface and the collection direction was about 45°. The collimated light was focused onto 230 μm diameter optical fiber connected to an Echelle spectrometer (Andor, Mechelle 5000) equipped with an ICCD camera (Andor, DH 734i), giving a broadband coverage from 220 to 850 nm with a resolution of 0.1 nm. The spectra were acquired with an 800 ns gate delay and 10 μs gate width in order to reduce the continuum emission. All the measurements were carried out in air at atmospheric pressure.

The studies were performed on 4 classes of standard geological cuttings: Quartz Sandstone, Brown Mudstone, Green Mudstone and Black Mudstone, which were from SINOPEC Shengli Geologging Company (SLGC). Each class contained 10 cutting samples. In order to meet the demand of real-time and on-line cuttings identification for

geological logging, no sample pretreatment was done in this work. We collected 50 single-shot spectra for each class, and representative LIBS spectra of the 4 different cutting classes are shown in Fig. 2.

3. Data analysis

The two chemometric methods PLS-DA and SVM used entire spectra as the input data, which were organized in a matrix containing the intensities at the different wavelengths (variables) in columns and the spectra (observations) in lines. The broadband LIBS spectrum consists of 24,041 intensity channels ranging from 220 nm to 850 nm, and each channel was used as an input variable. Then the dataset had 200 lines (4 cuttings \times 50 spectra) and 24,041 columns. To minimize the shot-to-shot variation effect of LIBS measurements, all spectra were normalized by the intensity of Ca II 393.4 nm because this line was found in all cutting samples with relatively high intensity. This way we performed the intra-class and inter-class normalization at the same time, and this pretreatment was shown to give the best results in our previous work [5]. After that, all variables were mean-centered, which was a common pretreatment for spectral data in the same unit. All the calculations in this work were performed using Matlab version 8.0 (Mathworks).

PLS-DA is a supervised multivariate least-squares analysis method used to classify samples [8]. It decomposes the spectra as linear combinations of principal components (PCs) that convey the major part of information contained in the global dataset. In PLS-DA, the predictor variables or latent variables (LVs) are generated from the input variables to maximize the variance between sample classes in the model, and it has been shown to be a more viable chemometric method than PCA and SIMCA [9]. The second chemometric method SVM is a relatively new nonlinear classification method which was proposed by Cortes and Vapnik in 1995 [17], and now has been successfully used in other spectroscopic techniques such as near-infrared spectroscopy [18] and fluorescence spectroscopy [19]. SVM constructs a set of hyperplanes in a high- or infinite-dimensional space, and a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called maximum-margin hyperplane). The principle of this method is illustrated in Fig. 3, in a simple case where there are only 2 classes [20]. The main difference between PLS-DA and

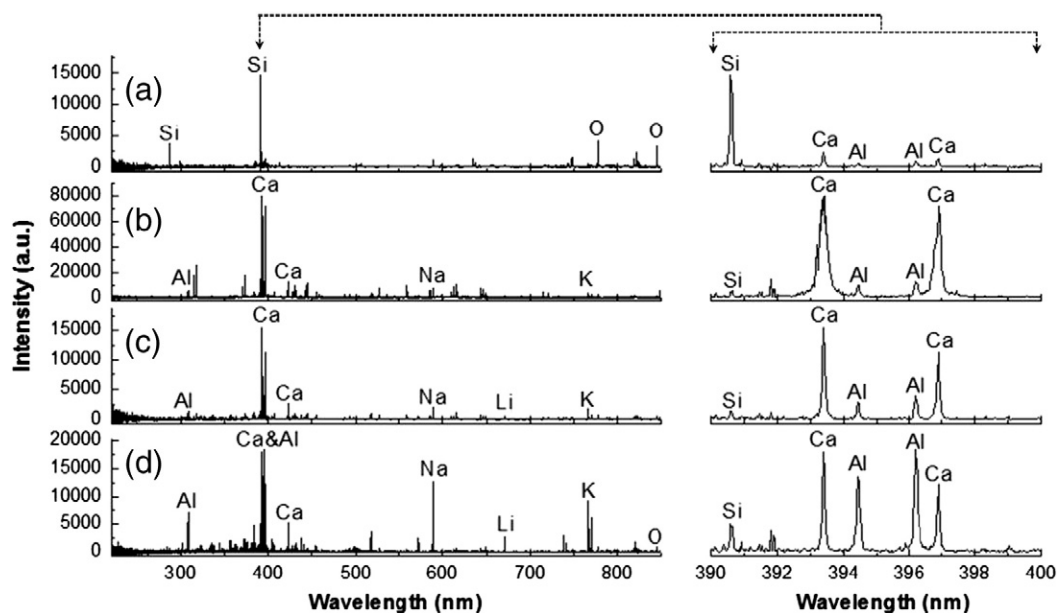


Fig. 2. Representative LIBS spectra of the cutting samples: (a) Quartz Sandstone, (b) Brown Mudstone, (c) Green Mudstone and (d) Black Mudstone. The observed emission lines from constituent elements Si, Al, Ca, Na, Li, K, and O have been identified. Inset of the spectra from 390 to 400 nm, which illustrate the cation compositions (Si, Al, Ca) of different cutting samples, is shown on the right side of the figure.

Download English Version:

<https://daneshyari.com/en/article/1239504>

Download Persian Version:

<https://daneshyari.com/article/1239504>

[Daneshyari.com](https://daneshyari.com)