Laser-induced breakdown spectroscopy applied to the characterization of rock by support vector machine combined with principal component analysis*

Hong-Xing Yang(杨洪星)[1], Hong-Bo Fu(付洪波)[1,2], Hua-Dong Wang(王华东)[1,2], Jun-Wei Jia(贾军伟)[1,2], Markus W Sigrist[3], and Feng-Zhong Dong(董凤忠)[1,2,†]

[1] *Anhui Provincial Key Laboratory of Photonic Devices and Materials, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei* 230031, *China*

[2] *University of Science and Technology of China, Hefei* 230026, *China*

[3] *ETH Zürich, Institute for Quantum Electronics, Otto-Stern-Weg* 1, *CH*-8093 *Zurich, Switzerland*

(Received 8 January 2016; revised manuscript received 1 March 2016; published online 20 April 2016)

Laser-induced breakdown spectroscopy (LIBS) is a versatile tool for both qualitative and quantitative analysis. In this paper, LIBS combined with principal component analysis (PCA) and support vector machine (SVM) is applied to rock analysis. Fourteen emission lines including Fe, Mg, Ca, Al, Si, and Ti are selected as analysis lines. A good accuracy (91.38% for the real rock) is achieved by using SVM to analyze the spectroscopic peak area data which are processed by PCA. It can not only reduce the noise and dimensionality which contributes to improving the efficiency of the program, but also solve the problem of linear inseparability by combining PCA and SVM. By this method, the ability of LIBS to classify rock is validated.

Keywords: laser-induced breakdown spectroscopy (LIBS), principal component analysis (PCA), support vector machine (SVM), lithology identification

PACS: 52.25.Kn, 52.50.Jm **DOI:** 10.1088/1674-1056/25/6/065201

1. Introduction

The laser-induced breakdown spectroscopy is a straightforward atomic emission spectroscopic technique that can provide rapid, multi-element detection with simple sample preparation. The laser-induced breakdown spectroscopy was first applied to the detection of hazardous gases and vapors in air at Los Alamos in the 1980s. Many researchers focused on the theory of laser-induced plasma (LIP),[1,2] electron density,[3,4] and laser–matter interaction which could improve the performance of LIBS.[5–8] Owing to its versatile advantages,[9] LIBS has been widely used in many fields from laboratory to practical applications such as industrial control,[10–13] environment protection,[14–17] agriculture,[18–20] medical science,[21,22] and archaeology.[23–25] Besides, LIBS has a trend of miniaturization, multiuse,[26] and using femtosecond laser to induce the plasma with some special characters.[27] Recently, qualitative analysis with chemometrics and pattern recognition has been conducted by many LIBS investigators and satisfactory results have been obtained by their work. In 2007, Sirven *et al.*[28] did many studies on the feasibility of rock identification at the surface of mars by remote laser-induced breakdown spectroscopy. They employed three chemometric methods: principal component analysis (PCA), soft independent modeling of class analogy (SIMCA), and partial least squares-discriminant analysis (PLS-DA) and obtained a best classification accuracy of 97.6%. In 2009, Yueh *et al.*[29] studied LIBS applied to the classification of biological samples by using 31 spectral lines of peak intensity, and they also gained good results by three methods: PLS-DA, hierarchical cluster analysis (HCA), and artificial neural network (ANN). In 2012, Cisewski *et al.*[30] built a support vector machine (SVM) model to evaluate the composition of suspect powders, particularly with respect to a possible content of Bacillus anthracis and obtained promising results.[30] De Lucia and Gottfried analyzed the explosive residues on organic substrates using LIBS data.[31] In China, many researchers, e.g., Yu *et al.*,[32] Tian *et al.*,[33] Wang *et al.*,[34] and Chen *et al.*[35] have contributed to LIBS for qualitative analyses through various methods such as PCA, PLS-DA, self-organizing mapping (SOM), and SVM. But the SVM combined with PCA has not been found to be applied to the LIBS data. In fact, the PCA is employed as a way of pretreating data and extracting features by many methods such as PLS-DA and SIMCA. This paper aims to classify rocks by LIBS data analyzed by SVM combined with principal component analysis.

*Project supported by the National Natural Science Foundation of China (Grant No. 11075184), the Knowledge Innovation Program of the Chinese Academy of Sciences (CAS) (Grant No. Y03RC21124), and the CAS President's International Fellowship Initiative Foundation (Grant No. 2015VMA007).

†Corresponding author. E-mail: fzdong@aiofm.ac.cn

© 2016 Chinese Physical Society and IOP Publishing Ltd

2. Theoretical analyses

2.1. LIBS theory

The LIBS process involves many complex but independent areas such as laser–matter interaction, laser ablation of material, optical and thermodynamic properties of hot and ionized gas, and plasma propagation in a background gas.[5] When the plasma is in local thermal equilibrium (LTE) and there is no self-absorption, the emission intensity is given by

$$I_{ij}=FA_{ij}h\nu_{ij}N\frac{g_i}{U(T)}\exp\left(-\frac{E_i}{k_{\mathrm{B}}T}\right), \tag{1}$$

where F is the experiment parameter, A_{ij} the transition probability for the spectral line from the energy level i to the level j, h the Planck constant, ν_{ij} the frequency of the light emitted from the energy level i to the level j, N the number of particles of all energy levels with respect to the content of the specific element, g_i the degeneracy of energy level i, $U(T)$ the distribution function of the particles representing the total states of the particles at all levels at plasma temperature T, E_i the excitation energy of level i, and k_{B} the Boltzmann constant. Equation (1) implies that the emission intensity depends linearly on the element concentration. However, in practice it is difficult to keep the plasma in LTE and self-absorption happens frequently.[36] Hence, the assumptions for Eq. (1) are difficult to realize and the element content and the spectrum intensity are not linear. Two typical ways are adapted to solve the problem. One is to employ the multivariate calibration methods,[37–40] and the other is to calculate the self-absorption and correct the line intensity.[41,42] But it can be assumed that there is some relationship between the intensity and the element concentration because the influence of self-absorption is constant for the same sample under identical experimental conditions. Instead, every selected line is a characteristic of the content of a specific element in the sample.

2.2. Principal component analysis (PCA)

PCA is often employed because it enables the decreasing of noise and dimensionality. The main ideal is that the raw data $\boldsymbol{X}$ can be decomposed into the product of two small matrices:

$$\boldsymbol{X}=\boldsymbol{T}\cdot\boldsymbol{L}^{\mathrm{T}}, \tag{2}$$

where the raw data $\boldsymbol{X}$ is an $n\times p$ matrix (n is the number of the samples, and p is the number of characteristics), $\boldsymbol{T}$ is the score matrix $n\times d$ (d is the number of principal components (PCs)), and $\boldsymbol{L}$ is the load matrix $p\times d$. The diagonal elements of $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}$ are called eigenvalues λ_i. In this way, the raw data are mapped into new dimensions. The new data $\boldsymbol{T}$ is a linear combination of the original data $\boldsymbol{X}$. The first PC can explain the highest variance of the raw data $\boldsymbol{X}$ because the PCA is based on maximal variance criterion. The second PC explains a little less variance than the first PC and the third again less than the second PC. Generally the first several PCs are sufficient when the sum of the PCs exceeds a threshold such as 90%, but it depends on the real situation. As a result, the dimension is reduced by abandoning the data which is dominated by noise.

2.3. SVM

The SVM has many applications in statistics, in particular for classification. The main idea of SVM is to find the hyperplane that can best distinguish the data by maximizing the margin between the closest points in each class.[43] Considering that the raw data may be nonlinear, the kernel function of the radial basis function (RBF) is chosen as:

$$K\left(x_i,x_j\right)=\exp\left(-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}\right), \tag{3}$$

where σ^2 is the RBF variable, which is set by experience initially and determined by cross validation in the end.

3. Experimental setup and data pretreatment

3.1. LIBS experimental system

The classic experimental setup is adopted to acquire the spectral data as shown in Fig. 1.

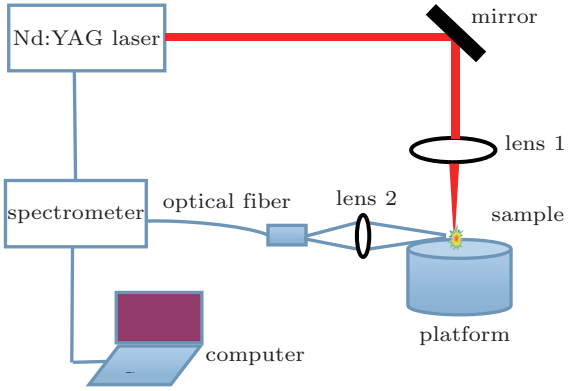


Fig. 1. (color online) Schematic experiment setup.

The Nd:YAG laser (Quantel model ULTRA50) is operated at the fundamental wavelength of 1064 nm. The laser pulse energy is fixed at 50 mJ, pulse width is 7ns, and the repetition frequency is 0.5 Hz. The laser is reflected by a 45° mirror first, and focused on the sample surface by an optical lens (lens 1) with 100-mm focal length. When the laser power density exceeds the ablation threshold of the material (typical 10^7 W/cm^2), a plasma is formed, emitting dazzling light which contains continuous background light and characteristic spectral lines from energy level transitions. The focused spot diameter is (0.75018 ± 0.06931) mm measured by Abbe comparator (Shanghai model 6W810115) for 5 different spots. Thus the laser power density can be calculated to be

1.26×10^9 W/cm^2. The emitting light is collected by a spectrometer (Avantes model AVS-DESKTOP-USB2) through an optical lens (lens 2) and a fiber. The spectrometer range is between 180 nm and 610 nm with a spectral resolution of about 0.05 nm. From the experiences of our group, the spectral integral time of the experiment is set to be 1.03 ms, and the delay time between the laser pulse and the collection of the emission is set to be 1.05 μs.

The experiment is performed with nine types of rock debris (Amphibolite, Gneiss, Limestone, Muddy Limestone, Shale, Quartz Sandstone, Basalt, Andesite and Granite). Figure 2 shows the approximate size of the rock debris sample which is pressed into a tablet under a pressure of 20 MPa for 5 min. Fourteen characteristic lines are picked corresponding to the emissions by Mg, Ca, Fe, Al, Si, and Ti, as listed in detail in Table 1. After that, the peak areas of the fourteen lines can be calculated. Because the intensity of Na emission line is higher than others and the spectrograph can become saturated easily, the Na lines are not included. From Fig. 3, it can be seen that all 14 emission lines do not have obvious self-absorption. But as is well known, the 14 emission lines are still influenced more or less by self-absorption, but it is hopeful to be partly overcome by the program since it is assumed that the degree of self-absorption of the same sample is the same. The 45 rock debris samples are chosen in the experiment. The first 5 laser pulses are employed to clean the sample surface and the subsequent 20 pulses are used to obtain an average spectrum.

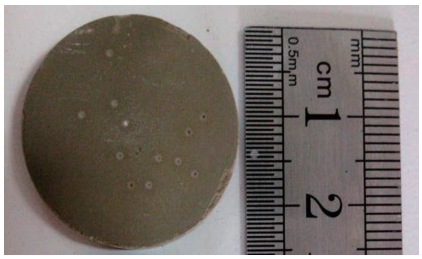


Fig. 2. Photograph of a rock debris sample.

Table 1. Selected emission lines for analysis of LIBS spectrum of rock debris.

Element	Analysis line/nm	Element	Analysis line/nm
Fe	275.57	Al	309.27
Fe	404.58	Al	394.38
Mg	280.27	Al	396.18
Mg	285.24	Ca	315.89
Mg	518.36	Ca	393.34
Si	288.15	Ca	396.85
Ti	334.94	Ca	585.74

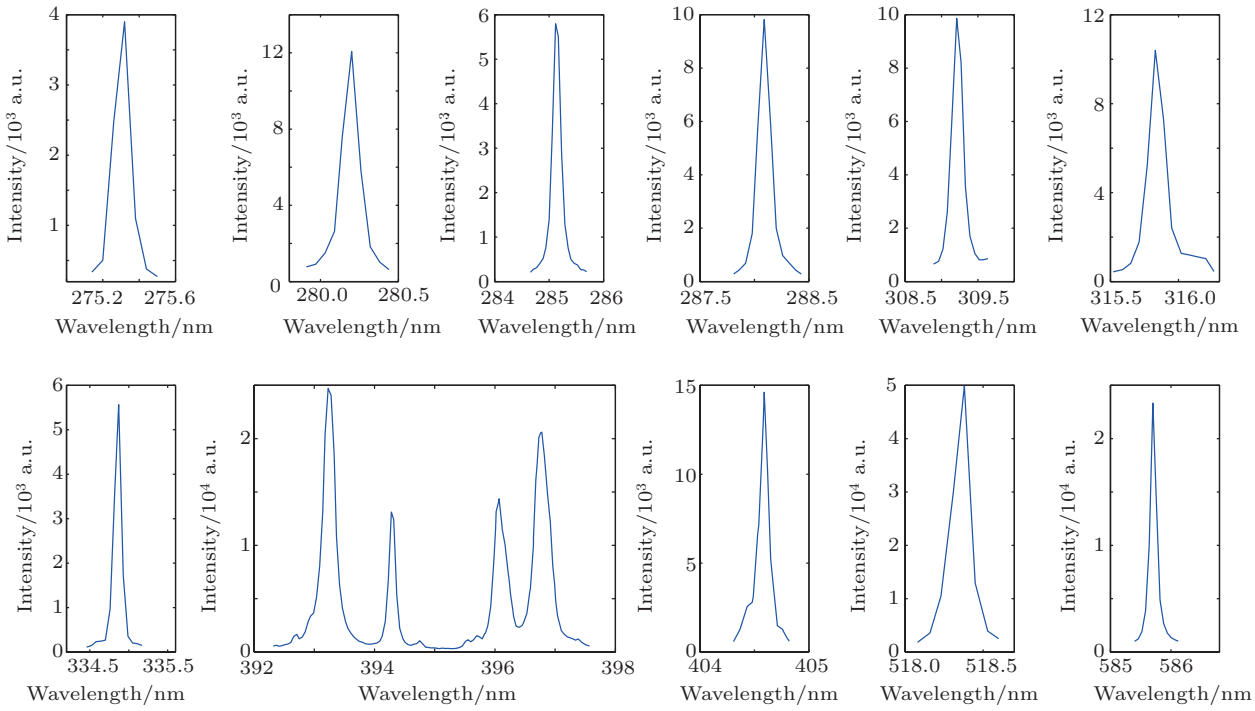


Fig. 3. (color online) Outlines of the 14 selected analytical lines.

3.2. Data pretreatment

The data selected from the total emission spectrum are the peak areas which are calculated from the spectra directly rather than the peak intensities of the 14 spectral lines. So the spectrometer does not need high spectral resolution. As an example, the intensity of the Ca line at 585.74 nm corresponds to 4956.60 a.u. (a.u. is short for arbitrary unit) as shown by the measurement in Fig. 4, while the Lorentz fitting yields 6280.05 a.u.

After the spectral peak areas are calculated, the data are normalized to reduce fluctuations caused by laser energy, irregularity of the sample surface and interference from outside. For each sample, the normalization is

$$X_i = \frac{D_i}{\sum_{1}^{14} D_i}, \tag{4}$$

where D_i is the peak area data calculated from the spectrum and X_i is the data after normalization. To build a better SVM training model, it is necessary to exclude singular samples. Lastly, the SVM yields the final classification after application by PCA.

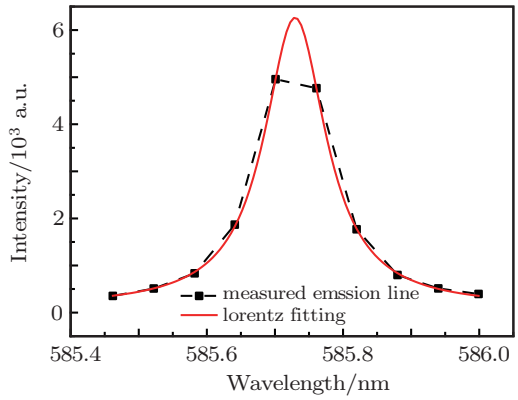


Fig. 4. (color online) Lorentz Fitting of Ca emission line at 585.74 nm.

4. Results and discussion

4.1. Result of PCA

PCA yields valuable information about the rock debris. At first it is necessary to determine the number of PCs. The first three PCs have explained over 90% of the variance of the original data and the fourth PC explained only 4.83% (see Table 2 for the details), so only the first three PCs are chosen. As a result, the dimensionality is reduced from 14 to 3.

Table 2. Rates of raw data variance explained by PCs.

PC number	1	2	3	4	5	6	7
Rate of variance explained	45.46	30.60	14.75	4.83	2.61	1.12	0.37
Sum	45.46	76.05	90.80	95.63	98.25	99.36	99.73
PC number	8	9	10	11	12	13	14
Rate of variance explained	0.16	0.04	0.04	0.02	0.01	0.00	0.00
Sum	99.89	99.94	99.97	99.99	100.00	100.00	100.00

As figure 5 implies, we can obtain a rough estimate about the classification ability on the basis of the first three PCs. Since each kind of rock debris has 5 samples, a visible saltation appears every 5 numbers at the boundaries of different kinds of rock debris. The PCs vary widely for different types of rock debris and change little within the same type. The first four kinds of rock debris can be easily distinguished by PC1. The sixth, eighth and ninth kinds can be set apart by PC2. The rest can be marked differently by PC3. The only minor defect is the fifth and ninth kind that may be mixed up. But the fifth and ninth kinds may be distinguished by PC4. However, it is possible that PC4 fails to get them apart for the following reasons: Firstly, it can be seen that the third PC has become smooth while it can explain 14.75% of the original variance and the fourth PC only 4.83%. Secondly, PCs are the linear combination of the original data. If the original data are nonlinear, the PCA is bound to fail no matter how many PCs are used.

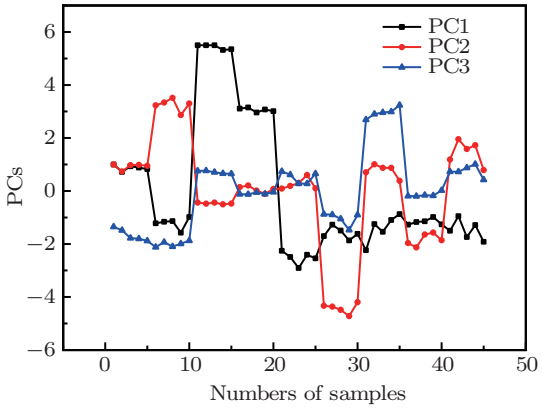


Fig. 5. (color online) Variations of the first three PCs with the number of samples. Each kind of rock is represented by 5 samples.

Figure 6 shows the scatters of the first three PCs of nine kinds of rock debris. The figure demonstrates that the nine kinds of rock debris are well distinguished and all five samples of each debris are gathered in a very small area. Only shale and granite (the fifth ■ and the ninth ▼) show a pair of overlapped points. Possible reasons of the overlapped points of shale and granite are below.

(i) Experiment fluctuation

Actually, LIP changes over time.[44] Besides that, the spectrum changes if the laser energy changes. However, since each pulsed laser shows some pulse power fluctuation, this phenomenon is within the experimental error. Thus if there are enough samples, it can be judged whether the result is within the experimental error by counting the overlapped points.

(ii) The spectral line of the Na element is not included in the analysis

As Table 3 shows, the largest difference between typical granite and shale is determined by the four elements Fe, Mg, Ca, and Na. The Fe, Mg, and Ca lines have been selected as analysis lines, but Na has not. Furthermore, the Na content of granite is about 10 times larger than that of shale, and represents the biggest among Fe, Mg, Ca, and Na. Thus if Na lines, such as 589.00 nm and 589.59 nm, are included for the analysis, PCA may be able to show better ability to differentiate between granite and shale.

Table 3. Element content (by weight percent) of typical granite and shale.

	Si (10^{-2})	Al (10^{-2})	Fe (10^{-2})	Mg (10^{-2})	Ca (10^{-2})	Na (10^{-2})	K (10^{-2})	Ti (10^{-6})
Granite	33.98	7.09	1.50	0.25	1.11	2.32	4.16	1720
Shale	27.64	9.96	5.31	1.21	0.43	0.26	3.45	3950

(iii) Granite and shale cannot be classified linearly

Since PCA just maps the data linearly, it is difficult to distinguish from granite and shale if they are not linearly separable. As a result, a new approach needs to be taken. There are many factors resulting in a linear inseparability such as self-absorption and inappropriate analysis lines. Hence the problem changes to how to select the most appropriate analysis lines which are free from self-absorption or how to correct intensity of analysis lines which have suffered self-absorption. Here the problem is turned into SVM.

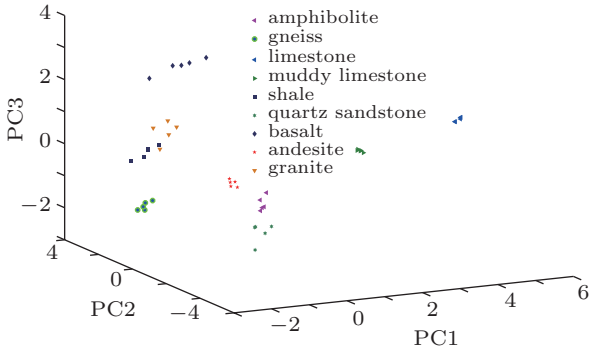


Fig. 6. (color online) Scatter of first three PCs of nine kinds of rock debris.

4.2. Results of SVM combined PCA

To overcome the defect of PCA, the SVM is employed to classify the nine rock debris based on the result of PCA. In addition, the RBF kernel function is chosen because of its ability to deal with nonlinear data. Five samples selected from each kind of rock debris are divided into three training samples and two test samples randomly. The first three PCs are chosen as the input data of SVM. To obtain a steady and reliable model, k-fold cross validation is operated to optimize the model. Two factors C and σ^2 vary from -1024 to $+1024$ in steps of 0.01 and from -100 to $+100$ in steps of 0.01 respectively. By this progress, the best optimized penalty coefficient $C = 1024$ and the RBF variable $\sigma^2 = 2.46$ are gained. The classification result is shown in Fig. 7.

It can be seen from Fig. 7 that all 18 test samples have been classified correctly with an accuracy of 100%, and also granite and shale are successfully classified. This good result can be explained by the following reasons.

i) The proper design of SVM combined with PCA

The noise reduction by SVM with the penalty coefficient C is limited, but PCA can compensate for that. PCA reduces the noise and dimensionality which improves the program efficiency. The most important is that the SVM has the ability to solve the linear inseparability. Thus PCA and SVM are mutually matched.

ii) The rock debris samples are homogeneous

Because the rock debris have been ground into powder, the samples are homogeneous as the photograph in Fig. 2 shows. But it is difficult to meet homogeneity during the excavation progress.

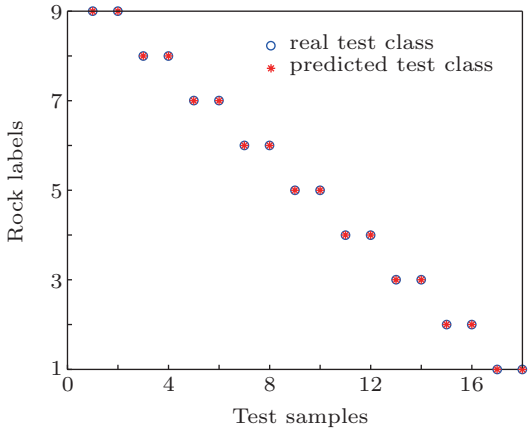


Fig. 7. (color online) Classification results of SVM combined with PCA for nine kinds of rock debris.

iii) Not enough samples have been examined

Hence, we investigate additional real rocks instead of rock debris by employing the peak areas of the same analysis lines with the same experimental setup and PCA combined with SVM. This time we chose 114 rocks of four kinds: sandstone, conglomerate, dolostone and mudstone. Figure 8 shows the results of 58 test samples. An accuracy of 91.38% (corresponding to 53 correct analyses out of 58 samples) is achieved when the penalty coefficient C is 1024 and the RBF variable σ^2 is 0.40. The accuracy of real rock samples is promising and is better than the SOM.[35] The five wrong classifications are mainly attributed to rock surface irregularity, inhomogeneous composition, and experimental fluctuations.

The whole program takes a few minutes (430.36 s), in which cross validation occupies much time (429.68 s, 99.84% of the total time). However, we should also realize that the model built by the training samples has been over trained. This can be derived from the penalty coefficient C of 1024 which is the highest among the numbers set by the program. The main reason is that the 14 selected emission lines cannot represent the biggest difference of the different kinds of rocks. The same situation happens during the identification of plastics.[32] The SVM is a supervised learning, so the training model is the key which determines the program performance. If the training

samples exhibit much noise, SVM must take a large penalty coefficient to classify them. Hence the noise reduction of the input data achieved by PCA keeps the training model more practical. The selection of the analytical emission lines plays a fundamental role in the training.

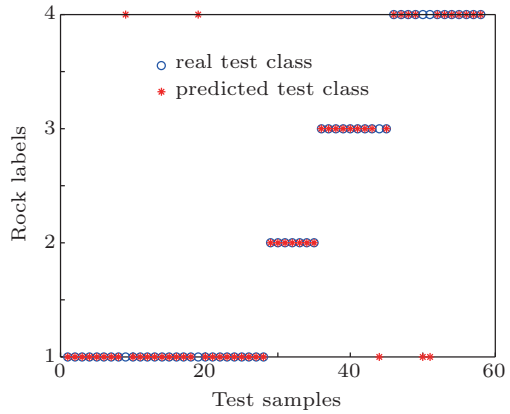


Fig. 8. (color online) Classification results of SVM combined with PCA for four real rock samples.

5. Conclusions

LIBS is demonstrated to be feasible for the analysis of rock during geological exploration due to its versatile advantages. The peak area instead of peak intensity is a suitable way to acquire useful data from LIBS spectra. For rock debris, the classification accuracy is 100%, while for real rock the accuracy decreases to 91.38%. LIBS yields rich and valuable spectral information. By combining PCA and SVM, it can not only reduce the noise and dimensionality which is helpful to improve the efficiency of the program, but also solve the problem of linear inseparability. This method can be adapted to classify the samples which have multivariate characters with not much noise. The crucial question is how to select the most appropriate emission lines for the analysis which can represent for the biggest difference among the different kinds of rocks in order to build a more practical model. Here we employ 14 analysis lines selected by the prior knowledge of the elements contained in the rock material.

References

[1] Cao Y, Liu X L, Xian W D, Sun S H, Sun M Z, Ding P J, Shi Y C, Liu Z Y and Hu B T 2015 *Chin. Phys. Lett.* **32** 035203
[2] Chen H, Lan H, Chen Z Q, Liu L N, Wu T, Zuo D L, Lu P X and Wang X B 2015 *Acta Phys. Sin.* **64** 075202 (in Chinese)
[3] Zhang Z M, Zhang B, Wu F J, Hong W, Teng J, He S K and Gu Y Q 2014 *Acta Phys. Sin.* **64** 105201 (in Chinese)
[4] Zhang L, Dong Q L, Wang S J, Sheng Z M and Zhang J 2010 *Chin. Phys. B* **19** 078701
[5] Yu J, Ma Q L, Motto-Ros V, Lei W Q, Wang X C and Bai X S 2012 *Front. Phys.* **7** 649
[6] Li Z C, Zheng J, Jiang X H, Wang Z B, Yang D, Zhang H, Li S W, Wang F, Peng X S and Yin Q 2011 *Chin. Phys. Lett.* **28** 125202
[7] Li Z C, Zheng J, Ding Y K, Yin Q, Jiang X H, Li S W, Guo L, Yang D, Wang Z B and Zhang H 2010 *Chin. Phys. B* **19** 125202
[8] Zhao Y, Zhu T, Wei M X, Xiong G, Song T M, Hu Z M, Huang C W, Shang W L, Yang G H and Zhang J Y 2012 *Chin. Phys. Lett.* **29** 085202
[9] Noll R 2012 *Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications* (Berlin: Springer) p. 7
[10] Lorenzen C J, Carlhoff C, Hahn U and Jogwich M 1992 *J. Anal. Atom. Spectrom.* **7** 1029
[11] Palanco S and Laserna J 2000 *J. Anal. Atom. Spectrom.* **15** 1321
[12] Kurihara M, Ikeda K, Izawa Y, Deguchi Y and Tarui H 2003 *Appl. Opt.* **42** 6159
[13] Noll R, Mönch I, Klein O and Lamott A 2005 *Spectrochim. Acta B* **60** 1070
[14] Tran M, Smith B W, Hahn D W and Winefoner J D 2001 *Appl. Spectrosc.* **55** 1455
[15] Hettinger B, Hohreiter V, Swingle M and Hahn D W 2006 *Appl. Spectrosc.* **60** 237
[16] Bossu M, Hao Z Q, Baudelet M, Yu J, Zhang Z and Zhang J 2007 *Chin. Phys. Lett.* **24** 3466
[17] Zhu G Z, Guo L B, Hao Z Q, Li C M, Shen M S, Li K H, Li X Y, Liu J G, Zeng X Y and Lu Y F 2015 *Acta Phys. Sin.* **64** 024212 (in Chinese)
[18] Sirven J B, Bousquet B, Canioni L, Sarger L, Tellier S, Potin-Gautier M and Le Hecho I 2006 *Anal. Bioanal. Chem.* **385** 256
[19] Bousquet B, Sirven J B and Canioni L 2007 *Spectrochim. Acta B* **62** 1582
[20] Senesi G, Dell'Aglio M, Gaudiuso R, De Giacomo A, Zaccone C, De Pascale O, Miano T and Capitelli M 2009 *Environmental Research* **109** 413
[21] Morel S, Leone N, Adam P and Amouroux J 2003 *Appl. Opt.* **42** 6184
[22] Corsi M, Cristoforetti G, Hidalgo M, Legnaioli S, Palleschi V, Salvetti A, Tognoni E and Vallebona C 2003 *Appl. Opt.* **42** 6133
[23] Burgio L, Melessanaki K, Doulgeridis M, Clark R J H and Anglos D 2001 *Spectrochim. Acta B* **56** 905
[24] Giakoumaki A, Melessanaki K and Anglos D 2007 *Anal. Bioanal. Chem.* **387** 749
[25] Kaszewska E A, Sylwestrzak M, Marczak J, Skrzeczanowski W, Iwanicka M, Szmit-Naud E, Anglos D and Targowski P 2013 *Appl. Spectrosc.* **67** 960
[26] Wang S H, Li Z L, Sun C L, Li Z W and Men Z W 2014 *Acta Phys. Sin.* **63** 205204 (in Chinese)
[27] Zhang N, Bao W X, Yang J H and Zhu X N 2013 *Chin. Phys. B* **22** 054209
[28] Sirven J B, Salle B, Mauchien P, Lacour J L, Maurice S and Manhes G 2007 *J. Anal. Atom. Spectrom.* **22** 1471
[29] Yueh F Y, Zheng H B, Singh J P and Burgess S 2009 *Spectrochim. Acta B* **64** 1059
[30] Cisewski J, Snyder E, Hannig J and Oudejans L 2012 *J. Chemometr.* **26** 143
[31] De Lucia F C and Gottfried J L 2012 *Appl. Opt.* **51** B83
[32] Yu Y, Hao Z Q, Li C M, Guo L B, Li K H, Zeng Q D, Li X Y, Ren Z and Zeng X Y 2013 *Acta Phys. Sin.* **62** 215201 (in Chinese)
[33] Tian Y, Wang Z N, Hou H M, Zhai X W, Ci X H and Zheng R E 2012 *Spectrosc. Spect. Anal.* **32** 2027
[34] Wang Q Q, Liu K and Zhao H 2012 *Chin. Phys. Lett.* **29** 044206
[35] Chen X L, Dong F Z, Tao G Q, Li Y J, She M J, Fu H B, Ni Z B and Wang J G 2013 *Chin. J. Lasers* **40** 1215001
[36] Cristoforetti G, De Giacomo A, Dell'Aglio M, Legnaioli S, Tognoni E, Palleschi V and Omenetto N 2010 *Spectrochim. Acta B* **65** 86
[37] Sun L X, Yu H B, Xin Y and Cong Z B 2010 *Spectrosc. Spect. Anal.* **30** 3186
[38] Stipe C B, Hensley B D, Boersema J L and Buckley S G 2010 *Appl. Spectrosc.* **64** 154
[39] Clegg S M, Sklute E, Dyar M D, Barefield J E and Wiens R C 2009 *Spectrochim. Acta B* **64** 79
[40] Laville S, Sabsabi M and Doucet F R 2007 *Spectrochim. Acta B* **62** 1557
[41] Bredice F, Borges F O, Sobral H, Villagran-Muniz M, Di Rocco H O, Cristoforetti G, Legnaioli S, Palleschi V, Pardini L, Salvetti A and Tognoni E 2006 *Spectrochim. Acta B* **61** 1294
[42] Bulajic D, Corsi M, Cristoforetti G, Legnaioli S, Palleschi V, Salvetti A and Tognoni E 2002 *Spectrochim. Acta B* **57** 339
[43] Cristianini N and Shawe-Taylor J 2000 *An introduction to support vector machines and other kernel-based learning methods* (Cambridge: Cambridge University Press) p. 93
[44] Liu Y F, Ding Y J, Peng Z M, Huang Y and Du Y J 2014 *Acta Phys. Sin.* **63** 205205 (in Chinese)