A comparative study of two data reduction methods for steel classification based on LIBS

Haiyang Kong\textsuperscript{1,2,3}, Lanxiang Sun\textsuperscript{1,3,a}, Jingtao Hu\textsuperscript{1,3}, Yong Xin\textsuperscript{1,3}, Zhibo Cong\textsuperscript{1,3}

\textsuperscript{1}Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
\textsuperscript{2}University of Chinese Academy of Sciences, Beijing 100049, China;
\textsuperscript{3}CAS Key Laboratory of Networked Control System, Shenyang 110016, China.
\textsuperscript{a}E-mail: sunlanxiang@sia.cn(corresponding author)

Keywords: Laser-Induced Breakdown Spectroscopy, Classification of steels, Artificial neural networks, Principal component analysis.

Abstract. Spectra of 27 steel samples were acquired by Laser-Induced Breakdown Spectroscopy (LIBS) for steel classification. Two methods were used to reduce dimensions: the first is to select characteristic lines of elements contained in the samples manually and the second is to do principal component analysis (PCA) of original spectra. Then the data after reducing dimensions was used as the input of artificial neural networks (ANN) to classify steel samples. The results show that, the better result can be achieved by selecting peak lines manually, but this solution needs much priori knowledge and wastes much time. The principal components (PCs) of original spectra were utilized as the input of artificial neural networks can also attain a good result nevertheless and this method can be developed into an automatic solution without any priori knowledge.

Introduction

An increasing number of garbage needs to be disposed with the development of industrialization of society. It is becoming increasingly important to recycle the waste materials in the garbage. The recycling of scrap steels is especially a matter of great significance\textsuperscript{[1-3]} The first step of recycling is to classify them and then to deal with them separately.

As a novel method of spectral analysis, Laser-Induced Breakdown Spectroscopy can achieve the purpose of quantitative or qualitative analysis of a sample through analyzing the plasma inspired by a high-energy laser. LIBS has gained rapid development as the sample can be analyzed by LIBS in situ and without contact\textsuperscript{[4-6]}

The spectra acquired by LIBS often contain hundreds of thousands of intensity lines due to the improvement of spectrometers. So it has increasingly become a difficult problem in LIBS analysis to extract useful information from the raw spectral data and reduce the dimensions of the original data.

Principal Component Analysis\textsuperscript{[7-9]} and Artificial Neural Networks\textsuperscript{[10-14]} were used to identify and classify materials widely as the common method of qualitative LIBS. Sun et al.\textsuperscript{[12]} trained an ANN using the characteristic lines selected manually to analyze the elements Mn and Si of steel samples and achieved a good result. Wang et al.\textsuperscript{[13]} selected 21 characteristic lines and trained an ANN network using the principal components of these lines. The network was used to classify plastic samples and an accuracy rate of 97.5\% was achieved.

We compared the two methods introduced above to reduce dimensions: the first is to select characteristic lines of elements contained in the samples manually and the second is to do principal component analysis of original spectra.

Experiments and Samples

Experimental setups. The schematic experimental setups of LIBS are shown in Fig. 1. A Nd-YAG Laser(BigSky, America) was used in the experiment. The output wavelength of the laser is 1064nm. The energy employed in this experiment was 55mJ and the focus point of lens was 1mm depth
under the sample surface. The laser frequency was set to 1 Hz. Spectra emitted by plasma were acquired by an Echelle spectrometer (LLA, Germany) with the gate delay 3µs and the integration time 5µs. Wavelength ranges from 200nm to 780nm and the resolution of spectrum is about 0.02nm (FWHM). Every spectrum contains 57144 pixels denoting intensities. Optimized parameters of the laser and spectrometer in previous studies in our laboratory were directly used.[12]

![Figure 1. Schematic experimental setups of LIBS.](image)

**Samples.** Twenty seven steel samples of four categories listed in Table 1 were investigated in the experiment. The first one was composed of 7 carbon steel samples. The second was 9 kinds of low alloy steels. The third was 5 high alloy steels. The last was composed of 6 stainless steels.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sample number</th>
<th>Sample name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-7</td>
<td>Carbon steels</td>
</tr>
<tr>
<td>2</td>
<td>8-16</td>
<td>Low alloy steels</td>
</tr>
<tr>
<td>3</td>
<td>17-21</td>
<td>High alloy steels</td>
</tr>
<tr>
<td>4</td>
<td>22-27</td>
<td>Stainless steels</td>
</tr>
</tbody>
</table>

**Spectra collection.** Four points of every sample were excited to acquire the spectra. The first 10 excitations were used to clean the surface of the sample and then the next 10 excitations were recorded to analyze the sample. So $10 \times 4 \times 27$ spectra were collected for 27 samples in total.

**Data sets:** Two data sets were used to verify the classification ability of the model. The first validation set is the average spectral data and it is composed of spectra collected from the last two points of samples. The second validation set is composed of the raw spectra from the last two points.

**Methods and discussions**

**Manually selecting characteristic lines.** The samples mainly contain nine elements, i.e. Mn, Si, Ni, Cr, V, Mo, Ti, Cu and Fe. Two characteristic lines were selected for every element and 18 lines were selected in all. The selected lines were listed in Table 2.

<table>
<thead>
<tr>
<th>Element</th>
<th>Mn</th>
<th>Si</th>
<th>Ni</th>
<th>Cr</th>
<th>V</th>
<th>Mo</th>
<th>Ti</th>
<th>Cu</th>
<th>Fe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lines(nm)</td>
<td>279.481</td>
<td>288.158</td>
<td>299.259</td>
<td>265.858</td>
<td>275.339</td>
<td>256.705</td>
<td>257.102</td>
<td>279.179</td>
<td>403.307</td>
</tr>
<tr>
<td>Lines(nm)</td>
<td>319.950</td>
<td>336.957</td>
<td>266.870</td>
<td>293.081</td>
<td>285.323</td>
<td>315.419</td>
<td>299.735</td>
<td>273.955</td>
<td></td>
</tr>
</tbody>
</table>

Then the corresponding intensities were trimmed from original spectra and an ANN network was trained using the intensities to classify the steels. The back propagation (BP) network was used widely in LIBS, so a three-layer BP network was used here. The number of neurons of input layer was the same as the input intensities. The neurons of hidden layer decided based on the experience formula $\sqrt{n+m}+a$ (Where $n$, $m$ are the number of neurons of the input and output layer respectively and $a$ is a constant between 0 and 10) was set to 8. One output neuron was utilized and the integers 1 to 4 was used to characterize four categories.

Figure 2 demonstrates the classification results using manually selecting lines as input.
As Fig. 2 shows, an excellent classification result was achieved using this ANN model. The accuracy rate of the average spectra by this model can reach 100% and the rate is still more than 99% for the non-average spectra. The manually selecting lines contain intensive and effective information and there are little noises in the data, so the outstanding result can be obtained. In addition, the complexity of the model is reduced as the amount of data was reduced after manually selecting characteristic lines.

**PCA of original spectra.** Although the good classification results can be obtained through manually selecting characteristic lines, this method needs spectroscopist with their priori knowledge to analyze spectra. It may waste much time to carefully compare and select, therefore the method will be restrained to a small range. An automatic classification method will promote this method to a great extent.

Principal component analysis was used to reduce the dimensions of the original spectra and a BP network was trained using the PCs to classify the steel samples. Six PCs that can explain more than 95% of the original spectra were used as the input of BP network. The number of neurons in the output layer is the same as that mentioned above. The neurons of hidden-layer was decided based on experience formula and the number is 7. The classification results of the BP network were shown in Fig. 3.

The classification results also achieved a good accuracy rate and the rate of two different data sets were both higher than 97%. So PCA can also extract the intensive information from the original spectra and remove the noises and get a good classification.
Comparative analysis. The classification results of two analysis methods were listed in Table 3.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>1st Validation Set</th>
<th>2nd Validation Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity lines</td>
<td>100</td>
<td>99.6</td>
</tr>
<tr>
<td>PCA of whole Spectra</td>
<td>97.8</td>
<td>97.6</td>
</tr>
</tbody>
</table>

As Table 3 shows, the excellent result can be obtained with the selected characteristic lines based on priori knowledge as input. However, the method need extensive experience of spectroscopist and wasted much time, so the application of this method was restrained. The ANN network trained using PCs of the whole original spectra also can achieve a comparative result to the results using characteristic lines although the accuracy rate is a little low. PCA data reduction method is a good analytical approach considering this method does not need human intervention and can be developed as an automatic solution.

Conclusions

We acquired spectra of 27 steel samples by LIBS and used two methods to reduce the dimensions of the data. The first one is manually selecting characteristic lines of spectra and the second is extract PCA of the whole spectra. Then the data after reduced dimensions were used as the input to train BP network to classify the steel samples respectively. The results show that, the BP network trained using characteristic lines can obtain a better classification, but it needs priori knowledge and experience of spectroscopist and wastes much time. Comparatively, the network trained using PCs of the whole spectra can also reach a good result. In addition, this method does not need the human interaction and can be developed to an automatic solution of qualitative LIBS.

Acknowledgements

The work was supported by the National High Technology Research and Development Program of China (863 Program) (No. 2012AA040608), the National Natural Science Foundation of China (No. 61004131) and the Development of Scientific Research Equipment Program of Chinese Academy of Sciences (No. YZ201247).

References

A Comparative Study of Two Data Reduction Methods for Steel Classification Based on LIBS

DOI References
http://dx.doi.org/10.1016/j.sab.2013.06.012
http://dx.doi.org/10.1039/c3ra44946g
http://dx.doi.org/10.1016/j.sab.2005.05.022
http://dx.doi.org/10.1002/app.33801
http://dx.doi.org/10.1366/10-06079
http://dx.doi.org/10.1007/s00216-006-0322-8
http://dx.doi.org/10.1007/s00339-008-4481-7
http://dx.doi.org/10.3788/AOS20103009.2757