

Identification of Materials Through SVM Classification of Their LIBS Spectra

Zuhaib Haider^a, Yusof Munajat^{a*}, Raja Ibrahim Kamarulzaman^a, Munaf Rashid^b

^aDepartment of Physics, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

^bFaculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding author: dryusofmunajat@gmail.com

Article history

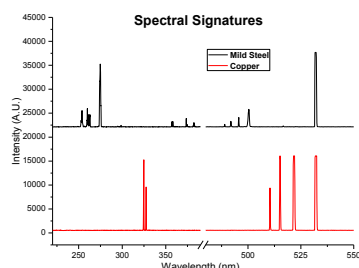
Received :18 March 2013

Received in revised form :

26 April 2013

Accepted :17 May 2013

Graphical abstract



LIBS Spectra of Mild Steel and Copper

Abstract

Laser Induced Breakdown Spectroscopy is a strong analytical method for qualitative studies and Support Vector Machines (SVM) is a powerful machine learning technique for pattern recognition and classification. In this paper we present an application of LIBS qualitative capability reinforced by SVM classification. Three different samples were ablated by an Nd:YAG laser and their spectra were recorded by Ocean Optics HR4000 spectrometer. These spectra possess signatures of the ablated materials. Sometimes these are visible to the naked eye while in many cases it is hard to decide about the presence of any pattern identifying a particular material. In addition variations are always found in the spectra obtained from laser induced ablation. In this situation a pattern recognition tool is very useful that sweep through the whole spectrum and record minor details. Here SVM serves the purpose. SVM classifiers were trained with distinct sets of spectra, belonging to specific materials, for classification. The results obtained from this preliminary experiment are encouraging and can lead us on positive grounds for the future work. This combination of tools can prove to be valuable for fast and automated identification and classification.

Keywords: Laser Induced Breakdown Spectroscopy (LIBS); machine learning; Support Vector Machines (SVMs); classification; accuracy

Abstrak

Laser Induced Breakdown Spectroscopy adalah satu kaedah analitikal untuk penentuan kualitatif manakala Support Vector Machines (SVM) ialah teknik pembelajaran mesin yang digunakan untuk pengenalanpastian dan pengklasifikasian corak. Dalam kajian ini, kami menggunakan aplikasi keupayaan kualitatif LIBS dan dibantu oleh pengklasifikasi SVM. Tiga sampel yang berlainan telah disingkirkan oleh laser Nd:YAG dan spektra yang terhasil direkodkan dengan menggunakan spektrometer Ocean Optics HR4000. Hasil spektra tersebut menunjukkan tanda-tanda bahan yang telah tersingkir. Kadangkala, tanda-tanda tersebut tidak dapat dilihat oleh mata kasar dan agak sukar untuk menentukan kehadiran sebarang corak untuk sesuatu bahan. Tambahan lagi, variasi sering berlaku pada spektra yang terhasil daripada laser aruhan yang disingkirkan. Dalam situasi ini, SVM adalah alat pengklasifikasi corak yang sangat berguna untuk merekod keseluruhan spektrum termasuk maklumat-maklumat yang terperinci mengenai sesuatu bahan. Pengklasifikasi SVM dilatih dengan pelbagai jenis set spektra yang merujuk kepada bahan yang tertentu sebelum pengklasifikasian yang sebenar. Daripada keputusan tahap ketepatan pengklasifikasian, pembaharuan SVM adalah satu kaedah untuk pengenalanpastian yang tepat dan pengklasifikasi spektra LIBS yang bagus. Penambahbaikan dari aspek kepantasan dan pengklasifikasi automatik mampu meningkatkan lagi keupayaan gabungan alat tersebut.

Kata kunci: Laser Induced Breakdown Spectroscopy (LIBS); pembelajaran mesin; Support Vector Machines (SVMs); pegklasifikasi; ketepatan

© 2013 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Laser Induced Breakdown Spectroscopy (LIBS) is a variant of Optical Emission Spectroscopy (OES). It is named so, because of using a powerful laser to generate plasma by producing breakdown on the sample surface [1]. The resultant plasma radiates out wide range of radiations with wavelengths and intensities specific to the elements and their concentrations in the plasma respectively. The elemental composition of the plasma is supposed to be the same as that of the sample. The emission spectrum obtained from the glowing plasma is thus representative of the original sample [2]. Therefore, two materials having different elemental compositions will generate different emission spectra. This fact can lead us to the identification of material species merely on the basis of their LIBS spectrum.

It is nearly impossible for a human eye to identify a material from its spectrum; however, it is possible with machine learning technique capable of pattern recognition. Supervised machine learning techniques are more suited for the purpose. A latest pattern identification tool based on supervised learning mechanism is known as Support Vector Machines (SVM), it is considered to be better than Artificial Neural Networks (ANNs) especially in multi-class classification [3]. Support vector machines use the concept of hyper-planes to find the decision boundaries and classify the problem by maximizing margin of the nearest instances to the decision boundaries called as support vectors [4]. SVM can handle both linear classification and more complex tasks of nonlinear classification as well it can handle multiclass problems. The idea used by svm is to plot the data into high-dimensional even infinite-dimensional feature space to

make the problem simpler. Kernel functions are used for this high-dimensional mapping. There are a variety of kernel functions available with SVM, the classification of the data depends upon the choice of kernel function used for data mapping [5].

Preprocessing of datasets is really important for better performance of classification tool. Data normalization is one of the common technique used for the data processing in order to bring the data values within a smaller range $\{-1,1\}$ or $\{0,1\}$. This helps in avoiding very large numbers to appear in the calculations. PCA (Principal Component Analysis) is another way of data processing; it helps in data reduction through feature selection giving a set of linearly uncorrelated variables called Principal Components [6], [7]. Since LIBS spectral data contains large amount of data for each instance, the data processing becomes a useful option to make the data handy. There are several examples where data preprocessing is employed as an essential part for LIBS data analysis [7], [8] and [9].

In this article, we briefly present the preliminary work for discrimination of materials on the basis of their LIBS spectra. It presents an experimental work on LIBS application for getting spectra and evaluation of SVM algorithm for their discrimination. Datasets (treated and untreated), training sets and kernel functions are evaluated for accuracy of the classification.

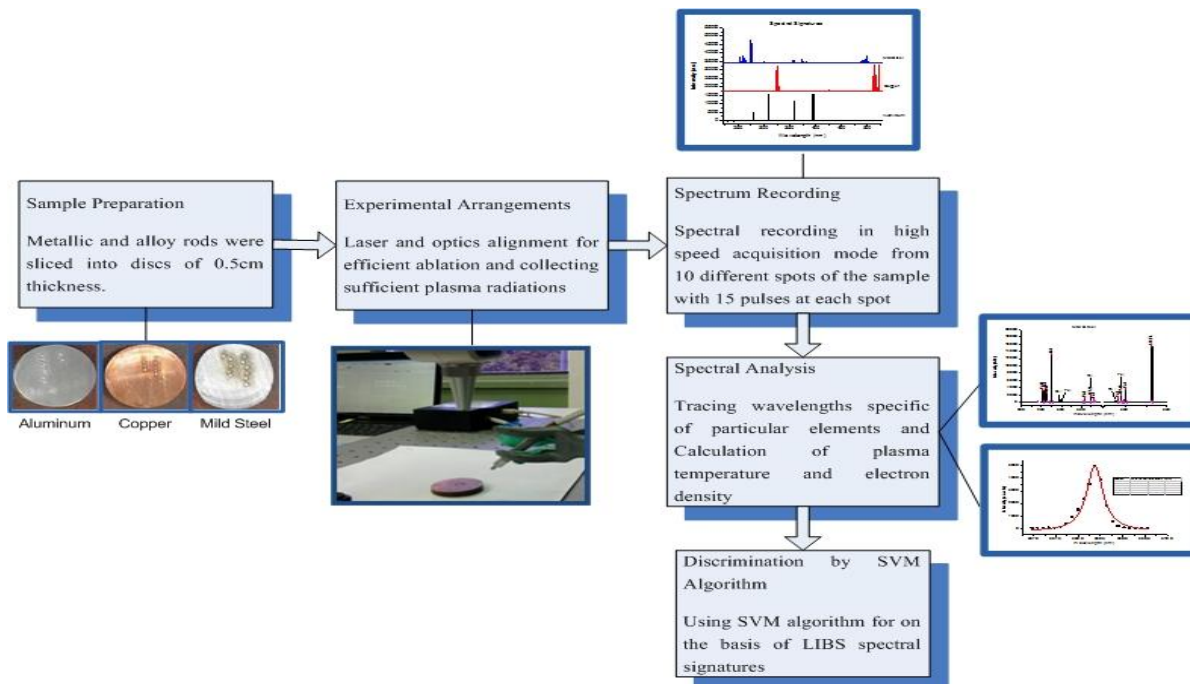


Figure 1 Methodology flow chart

2.0 EXPERIMENTAL PROCEDURES

2.1 Laser-Induced Breakdown Spectroscopy System

The LIBS system was used in its basic configuration. Nd:YAG laser with 6ns of pulse duration was operated at second harmonic frequency and energy of ~220mJ. Samples of Al, Cu and Mild Steel were used after cutting them into circular discs of 2.54cm and 5.8cm diameters with a thickness of 0.5cm. The sample was placed at tight focus of the laser beam. Ocean Optics HR4000 spectrometer was used for spectrum recording and was controlled by spectrasuite software provided by the manufacturer. For collecting and transmitting plasma radiations to the spectrometer a 600 μ m fiber optic cable was deployed.

2.2 LIBS Procedure

Samples were irradiated at 10 different spots keeping all the experimental parameters constant. Each spot was irradiated by 15 laser pulses at a frequency of 3Hz. From each set of laser pulses at one spot of the sample, spectra were continuously recorded for 5seconds employing high speed acquisition mode available through the software. In this way we collected more than 650 spectra from each spot of the sample, this provided us with the option of choosing spectra of interest. From the set of these recorded spectra, useful ones were separated and employed for subsequent studies.

2.3 Data Preprocessing

Data processing is of critical importance when large data sets are involved and calculations have to deal with big numbers. Here, we have tried data normalization technique for processing the data. Two types of data normalizations were performed and classification was tested on them. Three data sets are employed and classification performance of SVM classifiers is recorded with them. One data set contains the untreated data mentioned here as “Data Set 1” and is used as obtained, while the other two are normalized data sets, “Data Set 2” and “Data Set 3” normalized in the range of {0,1} and {-1,1} respectively. Variable results have obtained with these data sets under different circumstances.

2.4 SVM Classification

SVM classification is tested on spectral data of three different material species. Three different SVM classifiers (i.e., Default Kernel, Radial Basis Function (RBF) Kernel and Polynomial Kernel) were employed on aforementioned data sets. The training of a classifier is a crucial step and accuracy of classification depends very much on it. Here, we tried three training sets (namely “training set 1”, “training set 2” and “training set 3”) having different number of instances randomly chosen from the full data set by using ‘randperm()’ function, while the rest were used for testing. In full data set we had ten instances of each class of spectra (30 in total) which were then divided into training and testing data sets as just mentioned. For training, the number of instances in data sets were 16 (training set 1), 20 (training set 2) and 25 (training set 3), out of thirty the rest were used for testing the classifier’s efficiency. Figure 1 describes the whole methodology as was followed step-by-step. While flow chart of SVM algorithm is presented in Figure 2.

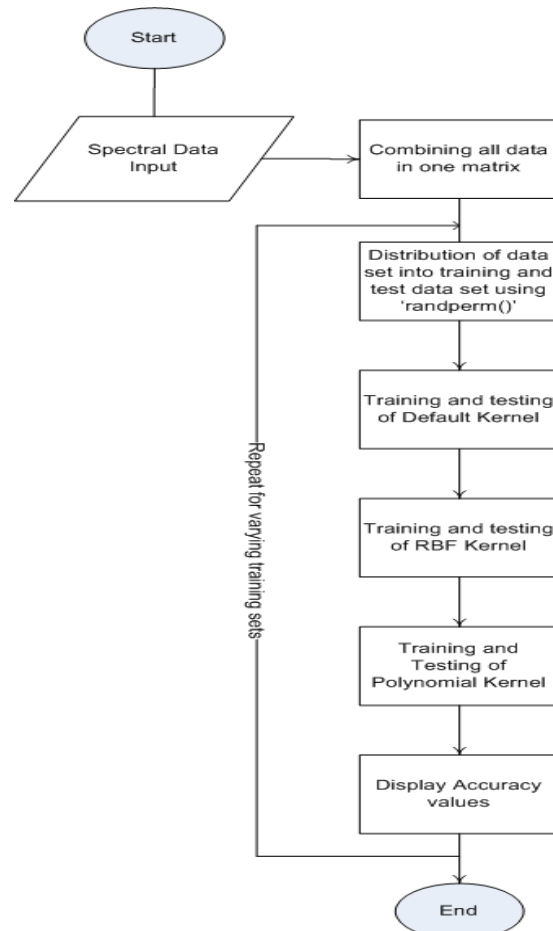


Figure 2 Flow chart of matlab algorithm for SVM classification

3.0 RESULTS AND DISCUSSION

The samples were ablated at 10 different spots by 15 laser pulses at each spot and spectra were recorded for each spot individually. Spectrometer was set to the high speed acquisition mode for continuous recording even when there was no laser (between two consecutive laser pulses). It kept on recording during the whole span of time adjusted by ourselves and it took 7ms on the average to record and save one spectrum. Exploring 10 different spots was to reduce the effect of possible heterogeneity (if any) in the sample composition and to make the data set more robust. In the first step the plasma was checked for localized thermodynamic equilibrium (LTE) and was found to be in LTE according to McWhirter’s criterion (explanation of McWhirter’s criterion is beyond the scope of this article, however, can be found in [1]). The spectra had no major differences because the samples were reasonably homogeneous but still we chose the best ones (i.e., having lower noise and stronger peaks) for better performance. Clear differences were observed in spectra of different samples due to different chemical compositions. Figure 3 shows the LIBS spectra belonging to aluminum, copper and mild steel behaving as their spectral signatures (these are the representative ones). Similar spectral instances constitute the data set but still with variations to make the data robust for better training of the SVM classifier.

Here, we have included three parameters for studying their effect on the classification accuracy of the SVM classifiers. (i)

Data sets (processed and unprocessed), (ii) No. of instances in training sets, (iii) Kernel Functions. In the following accuracy results obtained with each of these are described:

Unprocessed/Untreated and normalized data sets were employed for training of SVM classifiers/Kernels in the form of training sets. Figure 4(a) shows a bar graph of accuracies achieved with different training sets of each of the data sets for default kernel. Normalized data sets have contributed with better accuracies as compared to the unprocessed data set. Maximum accuracy obtained with training set 1 and 2 is for the normalized $\{-1,1\}$ data, whereas for training set 3 the situation turned out to be opposite, where we have obtained the least level of accuracy with the same data set.

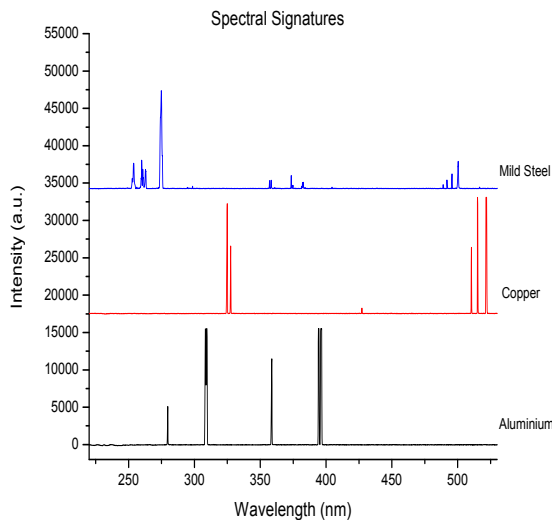


Figure 3 LIBS spectra of aluminum, copper and mild Steel

The bar graph Figure 4(b) depicts the maximum accuracy values that are achieved with three different kernels after training with three different training sets. It is observed that with training sets 1 & 2 all the three kernel functions generated same maximum accuracies i.e., 57% and 60% respectively. While for training set3 the maximum accuracies reached by kernel functions is variable where Polynomial kernel has outperformed others making 60% accurate classifications while default and RBF could produce only 40% accuracy.

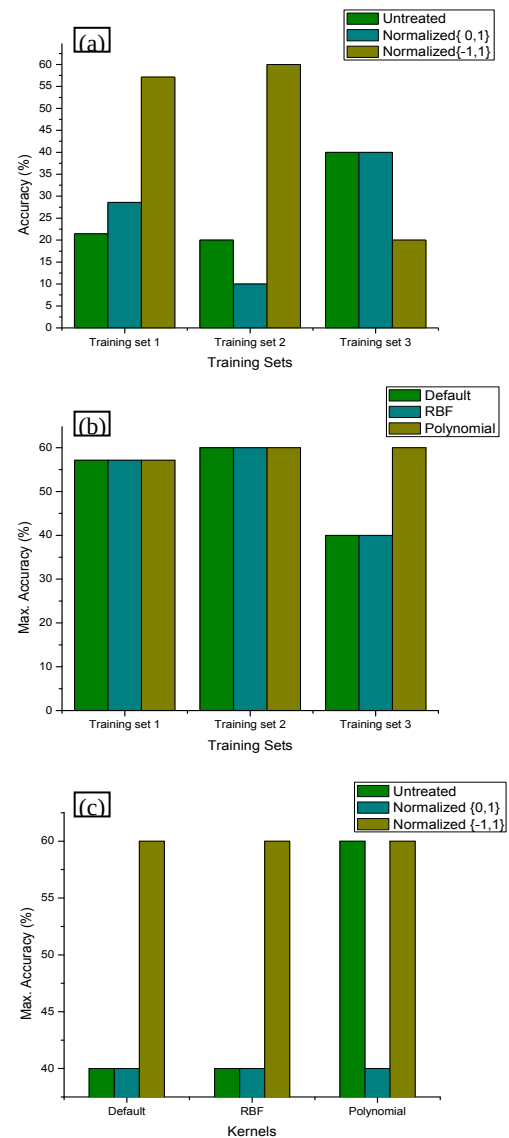


Figure 4 Classification accuracy graphs

Performance of kernel functions is visible through Figure 4(c). Maximum accuracies obtained with kernels with different data sets are presented. Maximum accuracy that these kernels could achieve with any data set is either 40% or 60%. It is clear that the performance of Polynomial kernel is better as for two out of three data sets it presented the peak accuracy of 60%. From all these results it is found that the best performances are delivered by the Polynomial kernel, the training set2 and the normalized $\{-1,1\}$ data set.

The maximum accuracy achieved with these kernels and data sets is 60%, while the lowest one is 10% that is achieved with default kernel trained at training set of the normalized $\{0,1\}$ data. The inconsistent variations in the accuracy levels of the classification, may be due to numerous reasons. The principal components scatter plot (Figure 5) for the data set that provided us with best results i.e., normalized $\{-1,1\}$ shows that the data is widely scattered and is mixed in several regions that makes it extremely difficult for classification. On the other hand, data may have been extrapolated due to very few training

instances as compared to the feature size of each instance, with which SVM classifiers could not be trained efficiently. Therefore, by the choice of least scattered data and selection of better training sets, we believe that classification problem can be far better solved.

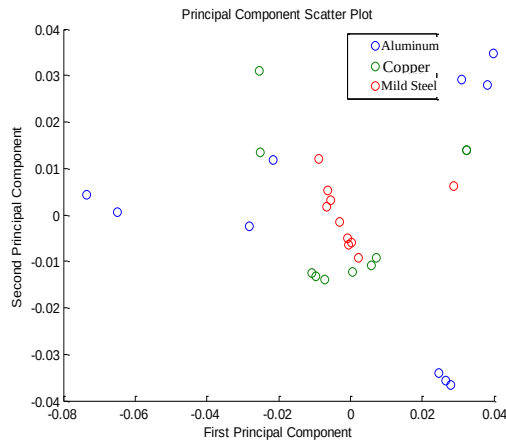


Figure 5 Principal component scatter plot

4.0 CONCLUSIONS

The results obtained from this preliminary work are not excellent but still are reasonable and encouraging. The peak accuracy achieved by each of the kernels is same i.e., 60%. However the variable trend in the classification accuracy mentions us about the weaknesses of the data set, improvement required in data processing and training of SVM classifiers. The training set 2, having 20 instances for training and 10 instances for testing of the kernel function, have shown better performance than others. The low levels of accuracy seem to be due to the selection of highly scattered data and extrapolation because of inefficient training of the classifiers by small training sets that lead to misclassifications. However, these results can

prove to be very useful for us in continuation of our work and will contribute to the improvement of the future work.

Acknowledgement

The authors would like to thank the Malaysian Ministry of Higher Education and Universiti Teknologi Malaysia for their financial support through ERGS grant R.J130000.7826.4L069 and International Doctoral Fellowship (IDF).

References

- [1] Singh, J. P. and S. N. Thakur. 2007. *Laser-Induced Breakdown Spectroscopy*. Oxford UK: Elsevier.
- [2] Noll, R. 2012. *Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications*. Berlin Heidelberg: Springer-Verlag.
- [3] Li, J. 2008. An Empirical Comparison between SVMs and ANNs for Speech Recognition. *Department of Computer Science, Rutgers University, USA*.
- [4] BURGESS, C. J. C. 1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*. 2: 121–167.
- [5] Abe, S. 2010. *Support Vector Machines for Pattern Classification*. London: Springer-Verlag
- [6] Remus, J. and K. S. Dunsin. 2012. Robust Validation of Pattern Classification Methods for Laser-induced Breakdown Spectroscopy. *Applied Optics*. 51(7): B49–B56.
- [7] Vance, T., N. Reljin, A. Lazarevic, D. Pokrajac, V. Kecman, N. Melikechi, A. Marcano, Y. Markushin and S. McDaniel. 2010. Classification of LIBS Protein Spectra Using Support Vector Machines and Adaptive Local Hyperplanes. *IEEE*. 1–7.
- [8] Dingari, N. C., I. Barman, A. K. Myakalwar, S. P. Tewari and M. K. Gundawar. 2012. Incorporation of Support Vector Machines in the LIBS Toolbox for Sensitive and Robust Classification Amidst Unexpected Sample and System Variability. *Analytical Chemistry*. 84: 2686–2694.
- [9] Cisewski, J., E. Snyder, J. Hannig and L. Oudejans. 2012. Support vector Machine Classification of Suspect Powders Using Laser-Induced Breakdown Spectroscopy (LIBS) spectral data. *Journal of Chemometrics*. 1–7.