Advances in Engineering Research, volume 127
3rd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2018)

Discussion on Machine Learning Data Processing Methods in Raman Spectrum

Wei He
The Third Research Institute of the Ministry of Public Security Shanghai 200004, China

***Abstract*—In this paper, exploiting mixed machine learning data processing methods in Raman spectrum, such as data organization, data preprocessing, data analysis, sample classification is discussed. Data experiments showed that the mixed machine learning data processing methods in Raman spectrum had good performance.**

Keywords—raman spectrum; machine learning; neural networks; sample classification

I. INTRODUCTION

Raman spectrum is based on inelastic scattering principle between light and material. It is modulated by the material structure and carries the characteristic spectrum of material. Raman spectrum analysis technique as an effective detection method has been widely used in many fields [1-3]. such as food, materials and environmental monitoring due to its advantages of non-destructive, rich information, no sample preparation etc. Handheld Raman spectrometer is widely used in the material identification of industrial production because of its advantages of easy to operate, compact construction, lightweight etc. However, during application, Raman spectrum noise and fluorescence are always the interfering facts to quality evaluation of Raman spectrum. Normally, In the past, much algorithm of data processing of removing background fluorescence, noise have been carried out during industry application [4-5]. Recently, machine learning and artificial intelligence is becoming hot in processing data of Raman spectrum [6-8]. In this paper, exploiting mixed machine learning data processing methods in Raman spectrum of handheld Raman spectroscopy are analyzed.

II. ANALYSIS METHODS OF MACHINE LEARNING DATA PROCESSING OF RAMAN SPECTRUM

The machine learning is an automatically processing data method, during Raman spectrum data processing, it includes procedure of data organization, data preprocessing, data analysis and sample classification.

A. Data Organization

Before processing data, all measurement data need to organized according to a certain format to achieve method easy and high efficiency. Usually we follow this data organization as shown in Form1: we organize the data to be processed according to the sample type (such as sample1~5), sample single measurement value (such as Sample1_001.txtt~sample1_006.txt), data arranged in rows according to wavenumber-spectral value in rows.

TABLE I. FORMATTING DATA

Sample type	Sample testing value	wavenumber-spectral value
Sample1	Sample1_001.txt	504.568, 1404
Sample2	Sample1_002.txt	507.523, 1358
Sample3	Sample1_003.txt	510.476, 1333
Sample4	Sample1_004.txt	513.426, 1307
Sample5	Sample1_005.txt	516.373, 1271

B. Data Preprocessing

Data preprocessing has varieties of functions, including data verification, bad data file elimination, data smoothing, noise filtering, artifact trace reduction and spectrum amplification etc.. Data verification and bad data file elimination are mainly used to verify data files and remove incomplete or damaged data files of sample. As shown in Figure I. Data smoothing and denoising mainly achieve the smoothing and noise reduction of sample data. The default preprocessing settings use a Median filter with a window width of 5 and a Wavelet filter with a progression of 5. Artifact trace reduction refers to the reduction of noise introduced by the optical lens itself. For example, lens artifact Raman spectrum, Sample Raman spectrum without artifact removal and Sample Raman spectrum after artifact removal are respectively shown in Figure II a, b and c.

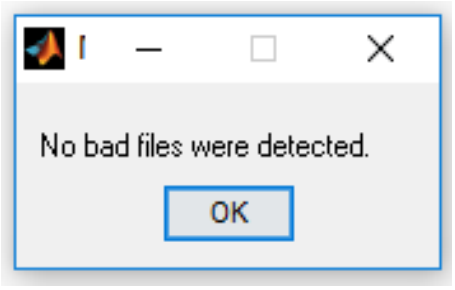
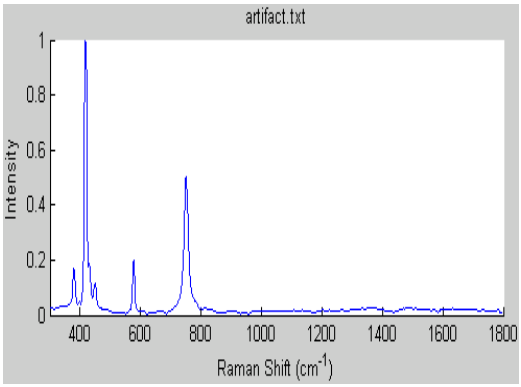
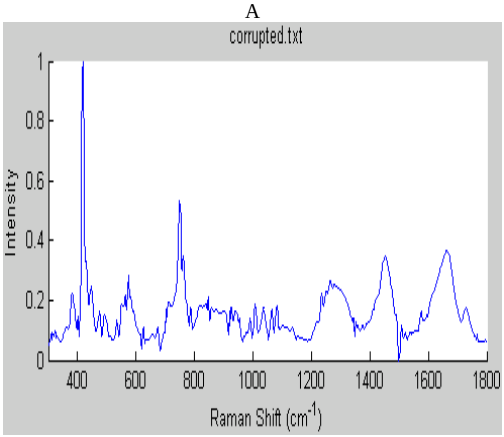


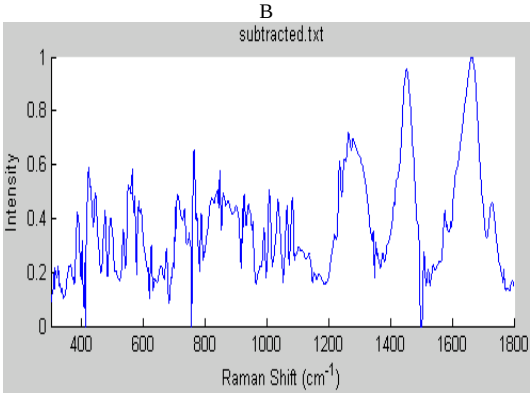
FIGURE I. DATA VERIFICATION



A



B



C

FIGURE II. A. LENS ARTIFACT RAMAN SPECTRUM; B. SAMPLE RAMAN SPECTRUM WITHOUT ARTIFACT REMOVAL; C SAMPLE RAMAN SPECTRUM AFTER ARTIFACT REMOVAL

C. *Data Analysis*

There is a variety of data analysis methods to analyze the Raman spectrum data. In this paper, we introduce tools of PCA (Principal Component Analysis) and NN (Neural Network).

Principal component analysis is a multivariate statistical method that examines the correlation between multiple variables. It studies how to reveal the internal structure of multiple variables by using a few principal components. That is, a few principal components are derived from the original variables, making them keep as much of the original variable information as possible and not related to each other. The usual mathematical treatment is to linearly combine the original P indicators as a new comprehensive indicator.

Neural network is a machine learning technology that simulates the neural network of human brain to implement artificial intelligence. We can use NN to train PCA data. Once a neural network is trained, it can be used to classify existing data. As shown in Figure III, we use NN to train PCA data By Matlab software to get the simulating result. There we present 2 groups Sample which are respectively cup, grape. Each group Sample has 10 pieces of data. We select one data to simulate with this method. The result is 100% correctly classified with MSE (Mean Square Error) is 0.3.42685e-010 after 33 epochs.

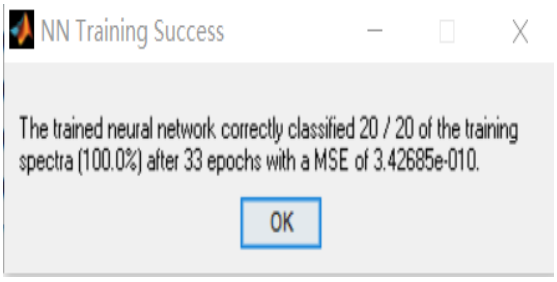


FIGURE III. THE NN TO TRAIN PCA DATA OF SAMPLE RAMAN SPECTRUM

D. *Sample Classification*

Using a trained classifier from Sample Raman spectrum after NN to train PCA data, a new Raman sample data file can be determined to which class it belongs to. We choose one data file randomly which is classified by NN. The Experimental results as shown in Figure IV, it belongs to cup group with 100% Correct rate.

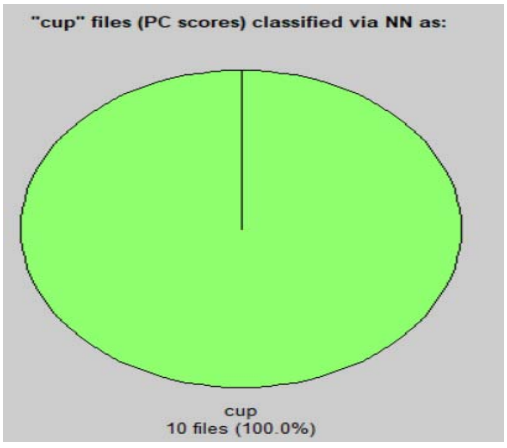


FIGURE IV. THE DATA FILE CLASSIFIED BY NN

III. CONCLUSION

The machine learning methods could make data processing Automatic, fast, and highly effective. With procedure of data organization, data preprocessing, data analysis and sample classification, experiments showed that the mixed machine learning data processing methods in Raman spectrum had good performance. In future, the handheld Raman spectroscopy with intelligent machine learning methods will

be considered as optimally effective detection method of data processing in Raman spectrum application.

Acknowledgment

COMMENTS: This study was funded by the 13th Five-Year National Key R&D Project “Flood Explosion Site Hazard Chemical Exploration Equipment and Accident Reconstruction Platform (Project ID: 2016YFC0801304)

References

[1] Mc Creery R L. Raman spectroscopy for chemical analysis [M]. New York: John Wiley & Sons, 2005, pp. 52-61

[2] Zhaoliang Gu; Weigen Chen; Yuxin Yun.. Silver nano-bulks surface-enhanced Raman spectroscopy used as rapid in-situ method for detection of furfural concentration in transformer oil , IEEE Transactions on Dielectrics and Electrical Insulation, Year: 2018, Volume: 25, Issue: 2 , pp.457 - 463

[3] Timothy H. Runcorn; Frederik G. Görlitz; Robert T. Murray., Visible Raman-Shifted Fiber Lasers for Biophotonic Applications, IEEE Journal of Selected Topics in Quantum Electronics, Year: 2018, Volume: 24, Issue: 3, Article Sequence Number: 1400208

[4] Bussian B M, Härdle W. Robust smoothing applied to white noise and single outlier contaminated Raman spectra [J]. Applied spectroscopy, 1984, 38(3), pp.309-313

[5] Rowlands C J, Elliott S R. Denoising of spectra with no user input: a spline‐smoothing algorithm[J]. Journal of Raman Spectroscopy, 2011, 42(3), pp.370-376

[6] K Xue, X Liu, J Yang, Raman Spectral Computer Data Processing System, Journal of University of Electronicence & Technology of China,1997

[7] The Raman Processing program is designed to process, analyze, and classify Raman spectra., http://cares.wayne.edu/rp

[8] IH Rodriguez, G Lopez-Reyes, DR Llanos, FR Perez, Automatic Raman Spectra Processing for Exomars, Springer Berlin Heidelberg , 2014, pp.127-130