CHEMICAL ENGINEERING TRANSACTIONS
VOL. 51, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Tichun Wang, Hongyang Zhang, Lei Tian
Copyright © 2016, AIDIC Servizi S.r.l.,
ISBN 978-88-95608-43-3; **ISSN** 2283-9216
DOI: 10.3303/CET1651190

Study on PCA-LDA for Fast Identifying the Type of Coal Mine Water with LIF Technology

Yong Yang[a,c*], Jing Li[b], Jianhua Yue[a], Li Zhao[b]

[a] School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, China
[b] School of Information Engineering, Southeast University, Nanjing, China
[c] School of Information and Electrical Engineering, College of Industrial Technology,Xuzhou, China
yongyang@cumt.edu.cn

Identifying the type of coal mine water is the foundation of coal mine hydrogeological research and has great significance for coal mine safe production. Considering the time consuming of the conventional method, we propose a new Laser Induced Fluorescence (LIF) combining with Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) method for coal source identification. Firstly, in order to obtain the valid bands from spectra results, LIF system is used to stimulate 405nm laser for the 400-800nm fluorescence spectra of tested water. Then, PCA is applied to reduce the dimension of spectral data. Finally, according to the different number of principal component with different pre-treatments the LDA identifications are implemented. Experiment results indicate that after valid bands selection, wavelet pre-processing, and PCA dimension reduction, the identification effect of spectra data with LDA method can be excellent as the number of principal component sets 6, and the correct recognition rate can reach to 100%. Thus, PCA-LDA algorithm with LIF tech is an effective identification method for quickly identification of the type of coal mine water.

1. Introduction

Mine water burst served as the second mine tremendous accident can cause huge casualties and economic losses in China. According to the statistics, during the period of 11th five-year plan there are over 5 times in a year on average and 506 death totally. Although many scholars have got improved achievement in the prevention and control of coal mine water burst, the overall performance is still insufficient. Also the problem of goaf water becomes more severe with the increasing of coal seam depth, combined with the complex mine hydrogeological condition and the indistinct water flowing fractured zone, the water mobility can hardly be carried out to monitor in effective way. Thus, according to the goaf water, surface water and the limestone water are likely to invade and mix in mine, it is important to select a proper method to identify the source of water accurately for inrush prevention.

Conventional identification methods of coal mine water source are QLT (water temperature and water level) Method (Liu et al., 2009), Representative Ion Method and Trace Element Method (Liu et al., 2011), etc. Yan Zhigang proposed a new model based on H support vector machine (SVM) for multiple source discrimination to identify the mine water inrush source (Yan et al., 2009); Wang Bingqiang presented a new method using system cluster analysis to detect the type of mine water source (Wang et al., 2015). Lu Jintao demonstrated Fisher linear discriminant function model and canonical discriminant function model based on the theory of Fisher discriminant analysis (FDA) to predict mine water type (Lu et al., 2012). However these methods need manual sampling and data analysis in laboratory, which causes large labour intensity, lagged monitored result and long cycle of data sampling. The precursor information of water inrush can hardly be achieved for the early warning.

Therefore, a new method is proposed to identify the type of inrushing water in this paper. Firstly, the fluorescence spectra of inrushing water are obtained by LIF technology for the valid bands from spectra results. Then pattern recognition is carried out for fluorescence spectra based on the theory of Linear Discriminant Analysis (LDA) for accurate coal source identification. LIF technology set the semiconductor laser with stable spectra and power as a laser light source and the micro optical fiber spectrometer as the receiver

Please cite this article as: Yan Y., Li J., Yue J.H., Zhao L., 2016, Study on pca-lda for fast identifying the type of coal mine water with lif technology, Chemical Engineering Transactions, 51, 1135-1140 DOI:10.3303/CET1651190

in on-line detection. The signal can be gathered by a stable and reliable fluorescent probe. Depending on the laser characteristics such as: orientation, cluster, stable spectra, narrow line width, the sensitivity of the sense line in fluorescence spectra can be greatly improved. So far laser induced fluorescence can be the new emerging technology in the field of vegetable oil identification(Wu et al., 2014), Toxicity of Heavy Metals in Water (Duan et al., 2013),monitoring of water quality(Li et al., 2006)etc. But the LIF technology combined with LDA algorithm for fast identification of coal mine water type has not been found in any relevant report.

2. Materials and Methods

2.1 Experimental Equipment

The laser induced fluorescence system (LIFS-405) made by Guang Dong Flag Electronics Co., Ltd is the main experimental equipment. The parameters set as: incident wavelength of laser is 405nm, incident power is 120mw, spectral range of detected fluorescence is 400-800nm, step length is 0.5nm, and spectral scan time is 1s/1000nm.Considering engineering application, we use the immersed fluorescent probe FPB-405-V3 (made by Guang Dong ke si kai Co.) to inspire the test water, instead of the conventional isolated laser excitation.

2.2 Experimental materials and spectral data acquisition

According to the relevant inrush accident research report, five kinds of water samples (ordovician limestone water, goaf water, alluvial water, sandstone water, limestone water) are collected from a mine at Huai Nan, An Hui province. Those samples are stored in dark sealed environment, 20 samples for each kind and 100 samples in all.

In the experiment, spectral data can be gathered from fluorescence excitation from 100 samples. To reduce the background light and human influence, instrument must work in a dark condition. Each sample measured 10 times, the arithmetic mean value is selected as the final spectral data.

Figure 1 indicates the significant differences among five kind's spectral data. Due to the lower resolution of the front and rear part of spectrum, the spectral data in the range of 420-670nm (500 data) band should be reserved for the dimension reduction and computation pressure alleviation. And the further lower dimension arithmetic mean of two adjacent data was calculated, where the dimension of spectral data in each group is reduced from the original 800 to 250.

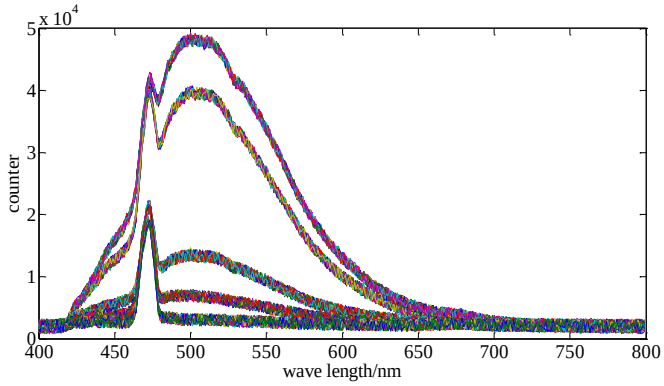


Figure1: Original fluorescence spectrum of water samples

2.3 The theory of experiment

LDA, a Pattern Recognition algorithm, projects the high-dimensional pattern sample into the optimal identification vector space for the classification information extraction, feature space dimension compression and pattern model both with maximum distance between classes and minimum distance within classes for the best reparability in the space. (Xie et al., 2010; Hua et al., 2008; Cao et al., 2008)

Suppose there are *m* samples,x_1, x_2, … , x_m, part of *c* classes. n_i Represents the number of samples which belongs to class *i*, so sample mean for class *i* is:

$$\overline{X_i} = \frac{1}{n_i} \sum_{x \in i} x \tag{1}$$

The overall sample means is:

$$\overline{X} = \frac{1}{m} \sum_{i=1}^{m} x \tag{2}$$

The scattering matrix between classes and the scattering matrix within classes from the macro view are defined by:

$$S_b = \sum_{i=1}^{c} n_i (\overline{X_i} - \overline{X})(\overline{X_i} - \overline{X})^T \tag{3}$$

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in i} (\overline{X_i} - x_k)(\overline{X_i} - x_k)^T \tag{4}$$

The aim of the LDA algorithm is to achieve the small coupling within classes and large coupling between classes. The identification criteria expression of LDA is defined by:

$$\Phi = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi} \tag{5}$$

Where φ represents *n*-dimensional column vector. For the best classification the maximum ratio between the square sum of different classes distance and the square sum of distance within classes should be achieved. Then, φ must meet the following conditions:

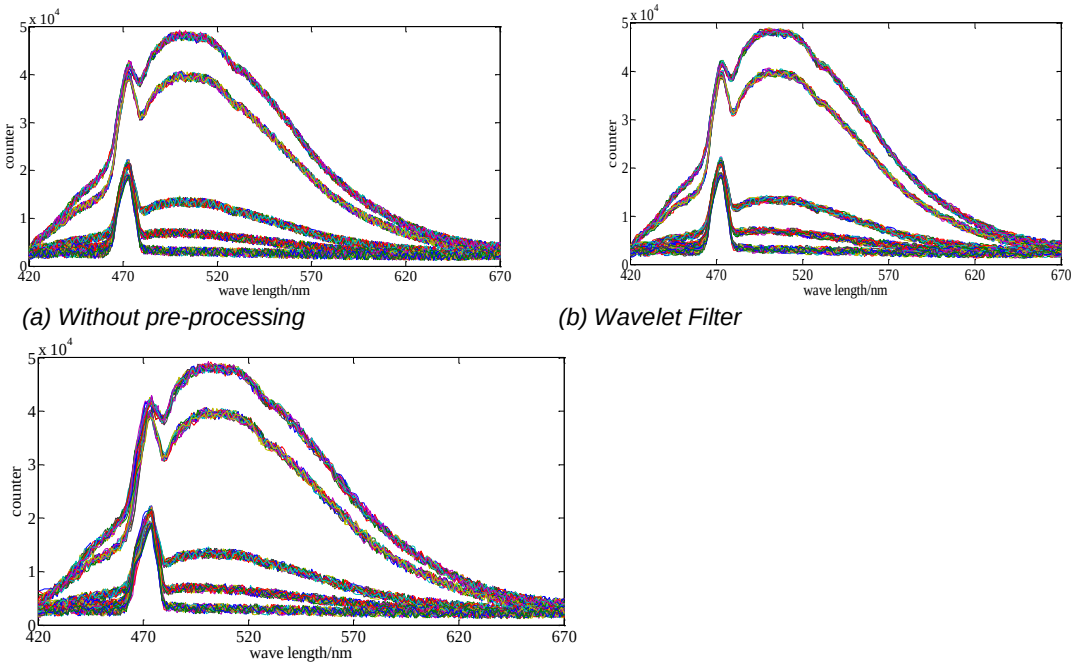
$$S_b \varphi = \lambda S_w \varphi \tag{6}$$

The column vectors φ are b (b≤ c-1) eigenvectors $S_w^{-1}S_b$ corresponding to the maximum eigenvalue. Here S_w must be non-singular, but due to the larger dimension of spectral data, the number of samples is much smaller than its dimension and the LDA can hardly work. Thus, it is important to reduce the dimension of sample.

3. Materials and Methods

3.1 2Preprocess for Spectral Data

In order to filter the amplitude of noise and improve resolution in spectral data, pre-processing on the fluorescence spectrum of water samples must be necessary. Median-Filter and wavelet algorithm can be carried out to conduct the best candidate. These experiment results of two kind methods will be compared to the original spectrum in the following Pattern Recognition step. The comparison result is shown in figure 2.



(a) Without pre-processing

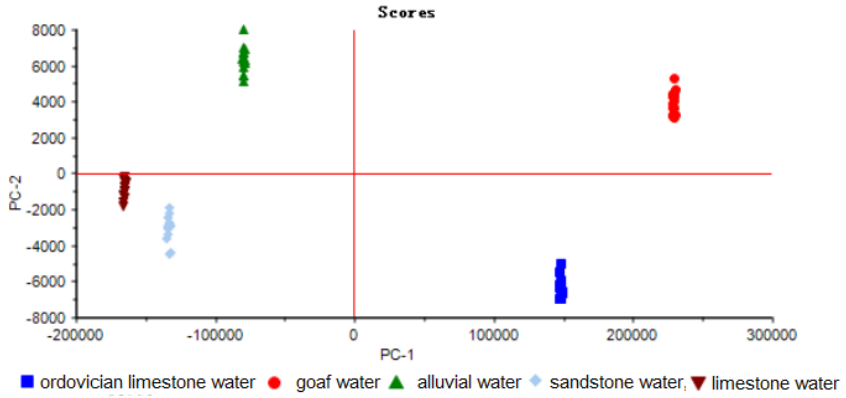
(b) Wavelet Filter

(c) Median-Filter

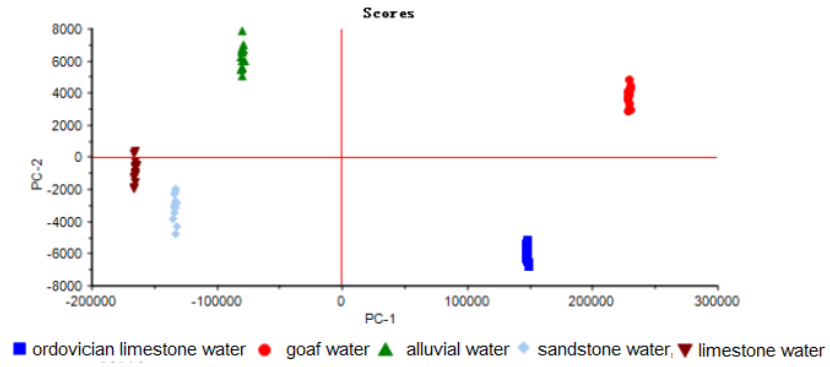
Figure 2: Different pre-processing results of spectrum

3.2 Dimension reduction with PCA

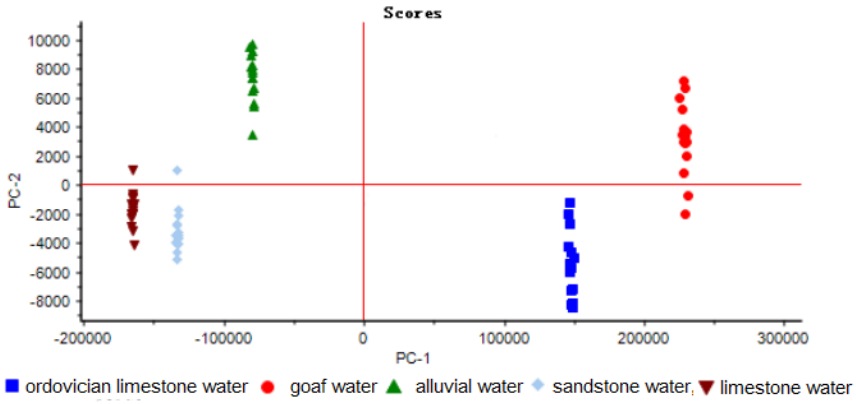
The experiment random selected 15 samples from the 5 kinds of water samples and 75 samples in total labelled as the training set, he rest 5 samples of the water and 25 samples in all labelled as test set. The training set is applied dimension reduction by PCA. The parameters set are the number of principal components as 10 and the fluorescence results as 3 classes mentioned in section 2.1. The scores of the first and second principal components are showed in figure 3. Compared with the hydro chemical analysis, the Ordovician limestone karstic water and the coal limestone water both belonging to limestone water are difficult to discriminate. The fluorescence spectrum of these two kinds of Limestone water can be recognized effectively as shown in Figure 3. The cluster results are obvious and the centres of the two clusters are far from each other.



(a) Without pre-processing



(b) Wavelet transforms



(c) Median-Filter

Figure 3: Score cluster results of with different pre-processing models

For the optimum number of principal components, Monte Calo Cross validation is used. The maximum number of principal components is set as 10, sampling rate is 0.9 and number of iterations is 1000. Table 1 shows the cross validation results of the pre-processed fluorescence, which indicate that the error rates of

cross validation (ER-CV) of the 3 kinds pre-process keep decreasing with the number of principal components increasing. When the principal components are set at 7, the ER-CV of the original spectrum and the MF spectrum both reach the minimum, 0.0012 and 0.0010 respectively. Also, the wavelet pre-processed spectrum reaches its minimum 0.0005 as the number of principal component is set at 6. Then the more principal components will not decrease the ER-CV.

Table 1: Results of Monte Carlo cross-validation simulation for different pre-processing

No. of Principal Components	Original	Wavelet Processed ER-CV	Median-Filter Processed
1	0.0657	0.0752	0.0610
2	0.0887	0.0637	0.0787
3	0.0742	0.0586	0.0682
4	0.0715	0.0503	0.0615
5	0.0564	0.0327	0.0524
6	0.0413	0.0005	0.0397
7	0.0012	0.0005	0.0010
8	0.0012	0.0005	0.0010
9	0.0012	0.0005	0.0010
10	0.0012	0.0005	0.0010

3.3 LDA classification

In order to explore the best number of principal component, LDA classification is set principal component as 6 and 7 respectively. The experiment was carried out in the test set with 25 samples and the results are shown in Table 2. The correct rate reaches 92% except 2 mistaken samples sandstone water from limestone water in original spectrum. After wavelet pre-processing, the correct rate of classification can reach 100% without miscarriage of justice. After MF pre-processing, the correct rate reaches 96% due to one sandstone water sample identified wrongly as Limestone water. When the number of principal component is 6, the correct rate can reach 88% since 2 water samples of sandstone is mistaken for Limestone water and 1 water sample of Limestone is mistaken for sandstone water in original spectrum after LDA classification. Using MF pre-processing, the correct rate also reaches 100% without miscarriage of justice. Using MF pre-processing, the correct rate reach 92% since 1 sandstone water sample is wrongly identified as Limestone water and 1 Limestone water sample is mistaken for water sandstone.

Table 2: Result of LDA classification of different pre-processing models

Accuracy	The original spectrum	wavelet	MF
LDA (the number of principal component 7)	92%(23/25)	25/25(100%)	24/25(96%)
LDA (the number of principal component 6)	88%(22/25)	25/25(100%)	23/25(92%)

After horizontal comparison, it shows that the result of the original spectrum classification is the worst with the same number of principal component, followed by the classification of MF spectra, the best classification is the spectrum with wavelet pre-processing whose correct rate reaches to 100%. By vertical comparison, the accuracy rates of original spectral classification and MF pre-processing are reduced with decrease of number of principal component. This is because the number of principal component contains the original spectrum information, so decrease of the number will cause loss of the spectral information and reduction in the correct rate.

Meanwhile, the correct rate using wavelet pre-processing reaches to 100% with number of principal component 6 and 7. It indicates that the number of principal component 6 stands for the enough spectral for LDA classification and the rest of the number of principal component belongs to redundant information for LDA classification. Thus, LIF technology combined with PCA-LDA method is an effective way to identify the coal mine water inrush source type.

4. Conclusion

In order to improve the time consuming of the conventional method, this paper propose a new LIF technology combining with PCA-LDA method for coal source identification. As is shown in the experiments, the number of principal component as 6, it can get a better result to identify the types of water source in coal mine using wavelet pre-processing and PCA for dimension reduction. In both the training set and the test set, the accuracy can reach 100% which proves the feasibility of PCA-LDA algorithm based on LIF technology for the quick identification of the coal mine water type. The work is further to expand the number of samples and explore the difference between the 5 kinds of mine water for the powerful theory evidence of real engineering applications.

Reference

Cao J., Zhang Y., Li J., Tang S., 2008, A Method of Adaptively Selecting Best LDA Model Based on Density, Chinese Journal Of Computers,(10):1780-178. DOI: 10.3321/j.issn:0254-4164.10.012.

Duan J., Liu W., Zhang Y., Zhao N., Wang Z., Yin G., Fang L., Liu J., 2013, Studies on Toxicity of Four Kinds of Heavy Metals in Water by Synchronous-Scan Fluorescence, Spectroscopy and Spectral Analysis,(05): 1262-1265, DOI: 10.3964/j.issn.1000-0593(2013)05-1262-04.

Hua J., Zhou Y., Liu T., 2008, Thermal Infrared Face Image Recognition Based on PCA and LDA, Pattern Recognition And Artificial Intelligence, (02):160-164, DOI: 10.3969/j.issn.1003-6059.2008.02.006.

Li H., Liu W., Zhang Y., Ding Z., Zhao N., Chen D., Liu J., 2006, Application of three-dimensional fluorescence spectrum for monitoring of water quality. Optical Technique, (01): 27-30, DOI: 10.3321/j.issn:1002-1582.2006.01.033.

Liu T.J., Zhang C.L., Qian J.Z., et al., 2011, Multivariate statistical analysis of trace elements in groundwater of old mining areas of Huainan coalfield, Journal of Hefei University of Technology, 34(01): 119, DOI: 10.3969/j.issn.1003-5060.2011.01.028.

Liu W.M., Gui H.R., Sun X.F., et al., 2009, QLT method used for determining water- inrush sources in Panji-Xiejiaji mines, China Coal, 27(5):31. DOI: 10.3969/j.issn.1006-530X.2001.05.011.

Lu J., Li X., Gong F., Wang X., Liu J., 2012, Recognizing of Mine Water Inrush Sources Based on Principal Components Analysis and Fisher Discrimination Analysis Method. China Safety Science Journal, (07): 109-115, DOI: 10.3969/j.issn.1003-3033.2012.07.018.

Wang B., Bai X., Wu Z., 2015, Prediction of mine water inrush sources based on cluster analysis of hydrogeo chemical features. Journal of Hebei University of Engineering (Natural Science Edition), (03):101-104, DOI: 10.3969/j.issn.1673-9469.2015.03.025.

Wu X.J., Zhao P., 2014, Application of Fluorescence Spectra and Parallel Factor Analysis in the Classification of Edible Vegetable Oils, Spectroscopy and Spectral Analysis,(08): 2137, DOI: 10.3964/j. issn. 1000-0593(2014) 08-2137-06.

Xie Y., 2010, LDA algorithm and its application to face recognition. Computer Engineering And Applications, (19): 189-192, DOI: 10.3778/j.issn.1002-8331.2010.19.055.

Yan Z.G., Bai H.B., 2009, MMH SUPPORT VECTOR MACHINES MODEL FOR RECOGNIZING MULTI-HEADSTREAM OF WATER INRUSH IN MINE,Chinese Journal of Rock Mechanics and Engineering,28(2): 324, DOI: 10.3321/j.issn:1000-6915.2009.02.015.