

# Simultaneous Differential Pulse Voltammetric Determination of Nitrophenol-type Compound by Artificial Intelligent and Machine Learning Methods

Ling Gao , Shouxin Ren

Department of Chemistry, Inner Mongolia University, Huhhot 010021, Inner Mongolia, China

**Abstract.** This paper suggests a novel method named DF-LS-SVM, which is based on least squares support vector machines (LS-SVM) regression combined with data fusion (DF) to enhance the ability to extract characteristic information and improve the quality of the regression. Simultaneous differential pulse voltammetric multicomponent determination of o-nitrophenol, m-nitrophenol and p-nitrophenol was conducted for the first time by using the proposed method. Data fusion is a technique that seamlessly integrates information from disparate sources to produce a single model or decision. The LS-SVM technique allows for learning a high-dimensional feature with fewer training data, and to reducing the computational complexity by only requiring the solution of a set of linear equations instead of a quadratic programming problem. Experimental results showed that the DF-LS-SVM method was successful for simultaneous multicomponent determination even when severe overlap of voltammograms existed. The DF-LS-SVM method is an attractive and promising hybrid approach that combines the best properties of the two techniques.

**Keywords:** artificial intelligence, machine learning methods, differential pulse voltammetric determination, nitrophenol-type compounds

## 1. Introduction

In the field of artificial intelligence and machine learning, raw data are often pre-processed to eliminate noise and irrelevant information prior to calibration. Wavelet transform (WT) [1, 2] and wavelet packet transform (WPT) have the ability to provide information in the time and frequency domain, so they can be used to convert data from their original domain into the wavelet domain, where the representation of a signal is sparse and it is easier to remove noise from the signal. Data measured for most processes are multiscale in nature, and since most data contain contributions at multiple scales, it is important to know which scales are the most important for forming a regression model with the highest prediction ability. Thus, developing multiscale approaches for data analysis and modelling relies on the extraction of various pieces of information from different wavelet scales, which is a challenging task. Unfortunately, in WT, multiscale methods are mentioned in a few papers [3, 4]. In fact, integrating the information from different wavelet scales is similar to processing large amounts of data from several sources by data fusion. Data fusion is one of the most important techniques in information science because it provides some vitally important perspectives. The research of data fusion in chemistry is at a preliminary stage, and the associated theory has rarely been explored in the field of information science. While data fusion has been applied in military affairs, robotics, remote sensing, image analysis and computer science, there are very few publications in the literature that address the utility of data fusion in chemistry [5].

Support vector machines (SVM) pioneered by Vapnik [6,7] are a kind of machine learning method based on modern statistical learning theory and have notable properties including the absence of local minima and high generalization ability. Suykens and his coworkers [8] introduced a modified version of SVM called least squares SVM (LS-SVM), which requires solving a set of linear equations instead of a quadratic programming problem, making it much easier and computationally simpler than SVM.

The aim of this work is to recommend a method named data fusion least squares support vector machines (DF-LS-SVM). This method combines the techniques of data fusion and multiscale wavelet transforms with LS-SVM regression to enhance the ability to extract characteristic information and improve the quality of the regression. The DF-LS-SVM method is a hybrid technique that combines the best properties of the two techniques and dramatically increases the problem-solving capacity. To the authors' best knowledge, this is

the first published work that describes the advantages of combining data fusion with a LS-SVM approach and the application of this method to perform simultaneous differential pulse voltammetric multicomponent analysis. This proposed method was applied to the simultaneous differential pulse voltammetric determination of o-nitrophenol, m-nitrophenol and p-nitrophenol.

## 2. Theory and Algorithms

### 2.1. Support vector machine regression

The objective of support vector machine regression (SVMR) is to find a regression function that relates the input data to the desired output properties:

$$y = wx + b \quad (1)$$

where  $w$  is the coefficient of regression function i.e. weight and  $b$  is the bias. These parameters are estimated by minimizing the following cost function:

$$R = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n L_{\varepsilon}(x_i, y_i) \quad (2)$$

$$L_{\varepsilon}(x, y) = \begin{cases} |x - y| - \varepsilon & |x - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where  $x$  is the desired response,  $y$  is the estimator output and  $\varepsilon$  is a prescribed parameter. In Eq. (2), the first term  $\frac{1}{2} \|w\|^2$  is used as a measurement of the model complexity. The second term is the so-called empirical error measured by the  $\varepsilon$ -insensitive loss function of  $L_{\varepsilon}(x_i, y_i)$ , which means that a subset of data points with absolute error not smaller than  $\varepsilon$ , the so-called support vectors, contributes to cost function  $R$ .  $C$  is a regularization constant that determines the trade-off between empirical error and the model complexity. The parameter  $\varepsilon$  and  $C$  must be defined by the user and they depend on problem and specified data. With the help of the Lagrange multiplier method and a quadratic programming algorithm, the Eq. (1) takes the following form:

$$y = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (4)$$

where  $\alpha_i$  and  $\alpha_i^*$  (with  $0 \leq \alpha_i, \alpha_i^* \leq C$ ) are Lagrange multipliers and the data points with nonzero  $\alpha_i$  and  $\alpha_i^*$  values are support vectors.  $K(x_i, x)$  represents the so-called kernel function, and any function that satisfies Mercer's condition can be used as the kernel function. In SVM, the radial basis function (RBF) and polynomial function are commonly used as kernel functions. The functions are defined in Eq. (5) and Eq. (6), respectively:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

$$K(x_i, x_j) = (x_i^T x_j + t)^d \quad (6)$$

where  $\sigma$  represents the kernel width of RBF,  $d$  denotes the degree of the polynomial kernel, and  $t$  represents the intercept. A remarkable feature of SVM is the application of the kernel function. By using the kernel functions, all necessary computations are performed directly in input space. Thus, instead of calculating a specific mapping for each dimension of the data, the problem only requires the selection of a proper kernel function and the optimization of its specific parameters.

### 2.2. Least squares support vector machines

LS-SVM is a modified of SVM proposed by Suykens [8] and has an important advantage of only requiring solving a set of linear equation instead of a quadratic programming problem. The LS-SVM method can perform both linear and nonlinear multivariate calibrations rather rapidly. During the training of the LS-SVM model, it is only necessary to solve a set of linear equations so the computational complexity is reduced.

### 2.3. Data fusion least squares support vector machine algorithm

Details about the DF-LS-SVM algorithm are presented below:

1. The whole set of spectra obtained from the standard mixture is used to build the experimental data matrix  $D$ . Before starting the DF-LS-SVM calculation, mean centering and data standardization are performed.

2. The matrix  $D$  is transformed to the WT domain by Mallat's algorithm. Due to its multiscale nature, the original signal (in the wavelength domain) is decomposed and separated into a set of multifrequency scales, each of which is treated as a separate source of information. In this paper a scale-dependent threshold method is performed. The wavelet coefficients of approximation and detail are reconstructed separately from a different scale denoised blocks. The reconstructed data in each scale, including the details at all levels, and the approximation at the coarsest level can be treated as a separate source of information. Some blocks that only contain noise and irrelevant information, such as  $Rd^1$ ,  $Rd^2$  and  $Rd^3$ , are removed. Other blocks which are most important for forming the calibration model, are retained and integrated into an augmented matrix. In essence, this is a data fusion process, which easily combined a large amount of data from different sources for obtaining a better regression model.

3. The augmented matrix that resulting from data fusion was used as input data to LS-SVM. The LS-SVM is a kernel machine, which is also applicable in regression application. The training of the LS-SVM model is performed with the augmented absorbance matrix and the concentration matrix as inputs. The concentrations of the predictive components can be calculated based on the trained LS-SVM model. The relative weight of the regression error  $\gamma$ , proper kernel function, and optimum kernel parameters are the crucial elements in LS-SVM. In this case, using the RBF kernel,  $\gamma$  and kernel width  $\sigma^2$  parameters are selected by trial and error. Currently, there is no theoretical guidance to select the optimal values of  $\gamma$  and  $\sigma^2$ , which depend on the specific problem and the data themselves.

## 3. Experimental Section

A BAS-100A Electrochemical Analyzer (Bioanalytical Systems, Inc.) was used for all experiments. A PAR 303A static mercury drop electrode (Princeton Applied Research) was used as a working electrode, a Pt wire as a counter electrode and an Ag/AgCl/ saturated-KCl as a reference electrode. Differential pulse voltammograms (DPV) were recorded between  $-200$  mV and  $-500$  mV at 4.0 mV intervals. The whole set of voltammograms obtained in 16 standard mixtures was used to build up the matrix  $D$ . Using the same procedures, a  $D_u$  matrix for nine synthetic test samples was built up.

## 4. Results and Discussion

### 4.1. Wavelet transform and wavelet multiscale properties.

In order to visually inspect the multiscale property of WT, the scales ( $a^7$ ,  $d^7$ ,  $d^6$ ,  $d^5$ ,  $d^4$ ,  $d^3$ ,  $d^2$  and  $d^1$ ) were converted back to their original domain by Mallat's algorithm to retain the original length of the signal. The reconstructed components ( $Ra^7$ ,  $Rd^7$ ,  $Rd^6$ ,  $Rd^5$ ,  $Rd^4$ ,  $Rd^3$ ,  $Rd^2$  and  $Rd^1$ ) are shown in Figure 1. These reconstructed components are mutually orthogonal and contribute to the original signal from different scales having difference frequencies. From Figure 1, it is obvious that  $Rd^1$ ,  $Rd^2$  and  $Rd^3$  centered in the high frequency ranges looked like noise and should be eliminated. Other scales ( $Rd^7$ ,  $Rd^6$ ,  $Rd^5$  and  $Rd^4$ ) contain pertinent information and should be retained, because analytical signals usually are found in the low frequency ranges. Scale  $Ra^7$  is the approximation of the coarsest level and concentrated in the lowest frequency ranges may contain some backgrounds and drift, but the useful information may be embedded in

them. To eliminate  $Ra^7$  is in danger of losing analytical information. Thus, the scale ( $Ra^7$ ) is selected and added the augmentation data matrix.

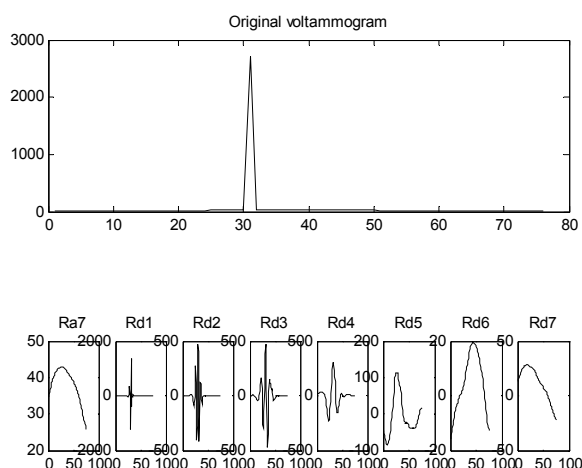


Figure 1. Plots of the visual inspection of the multiscale property of WT

## 4.2. DF-LS-SVM

In order to optimize the DF-LS-SVM method, five parameters, i.e. relative weight of regression error ( $\gamma$ ), kernel width  $\sigma$  of RBF, wavelet functions, decomposition level, and thresholding methods need to be optimized by trial and error. As the results of the optimization,  $\gamma=2000$ ,  $\sigma^2=400$ , Daubechies 4,  $L = 7$ , and the HYBRID thresholding were selected as optimal parameters.

The DF-LS-SVM method relies on the idea of fusing multiple data sources and performing LS-SVM on this fused matrix in order to produce a global estimate of the concentrations of different compounds. In this case, some data sources located in the high-frequency ranges such as  $Rd^1$ ,  $Rd^2$  and  $Rd^3$  only contain noise and irrelevant information and were removed before the data fusion operation. The rejection of uninformative input variables from the model may improve the quality of regression. Thus, a rough idea of which data sources are most important for forming a LS-SVM model with the highest prediction ability can be obtained. After reconstruction the important data sources were converted to the original domain and were combined by data fusion techniques. The augmented matrix was applied as input data to LS-SVM method. In order to evaluate the advantages of data fusion, a comparison of DF-LS-SVM regression and single LS-SVM regression from different data sources was performed. Five kinds of single LS-SVM methods from different data sources including  $Rd^4$ ,  $Rd^5$ ,  $Rd^6$ ,  $Rd^7$  and  $Ra^7$  were designed. The predictive parameter RSEP for total compounds of the same test set was computed. It can be seen that the DF-LS-SVM method performs better than all single LS-SVM methods. DF-LS-SVM regression with the advantages of data fusion obviously produced superior effects and this method is a challenge for multivariate regression.

Using the DF-LS-SVM method, the concentrations of o-nitrophenol, m-nitrophenol and p-nitrophenol for the test set were calculated. The experimental results showed that the RSEP for total compound was 3.9 %.

## 4.3. A comparison of PLS, WT-PLS, LS-SVM and DF-LS-SVM

In order to evaluate the DF-LS-SVM method, four methods were tested in this study with a set of synthetic, unknown samples. The RSEP for the four methods are displayed in Table 1. The RSEP for total compounds with DF-LS-SVM, LS-SVM, WT-PLS and PLS were 3.9%, 5.6%, 5.0%, and 7.6%, respectively. The RSEP values indicate that the DF-LS-SVM, LS-SVM and WT-PLS perform better than PLS methods, and the DF-LS-SVM method has the best performance among the four methods. The DF-LS-SVM method is a hybrid technique that combines the best properties of DF and LS-SVM methods and dramatically increases the problem-solving capacity. The results demonstrate that the DF-LS-SVM method performed well and is a promising technique.

## 5. Conclusions

A method named DF–LS-SVM that is based on LS-SVM regression combined with data fusion was developed for multicomponent differential pulse voltammetric determinations. In this case, the DF–LS-SVM method was proven to be successful even when severe signal overlap was present, and it performed better than the LS-SVM, WT-PLS, and PLS methods. However, the method is still new and many of its aspects are yet to be fully tested. Future work is needed to refine and improve this approach.

TABLE I RSEP VALUES FOR O-NITROPHENOL, M-NITROPHENOL AND P-NITROPHENOL SYSTEM BY THE THREE METHODS

Method	RSEP (%)			
	I <sup>a</sup>	II <sup>b</sup>	III <sup>c</sup>	Total compounds
DF-LS-SVM	4.1	3.7	3.9	3.9
LS-SVM	7.3	5.0	4.0	5.6
WT-PLS	8.1	2.4	2.3	5.0
PLS	9.9	5.6	4.1	7.6

<sup>a</sup>: o-nitrophenol ; <sup>b</sup>:m-nitrophenol; <sup>c</sup>: p – nitrophenol

## Acknowledgments

The authors would like to thank National Natural Science Foundation of China (21067006 and 60762003) and Natural Science Foundation of Inner Mongolia (2009MS 0209) for financial support of this project.

## References

- [1] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets,” *IEEE Trans. Inform. Theory*, 1992, **38** (2): 617-643.
- [2] S. X. Ren and L. Gao, “Simultaneous quantitative analysis of overlapping spectrophotometric signals using wavelet multiresolution analysis and partial least squares,” *Talanta*, 2000, **50** (6):1163-1173.
- [3] Alsberg, B. K. “Multiscale cluster analysis,” *Anal. Chem.*, 1999, **71**: 3092-3100.
- [4] Z. Ahmad,; J. Zhang, “Combination of multiple neural networks using data fusion techniques for enhanced nonlinear process modeling,” *Computers and Chem. Engineering* 2005, **30**: 295-308.
- [5] K. H. Ruhm, “Sensor fusion and data fusion-mapping and reconstruction,” *Measurement*, 2007, **40**: 145-157.
- [6] V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag, New York, 1995.
- [7] V. N. Vapnik, “An overview of statistical learning theory”, *IEEE Trans. Neural Networks*, 1999, **10**: 988-999.
- [8] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least-squares support Machines*, World Scientific: Singapore, 2002.