

Non-linear carbon dioxide determination using infrared gas sensors and neural networks with Bayesian regularization

King-Tong Lau, Weimin Guo[#], Breda Kiernan, Conor Slater, Dermot Diamond*

Adaptive Information Cluster, National Centre for Sensor Research, Dublin City University, School of Chemical Sciences, Glasnevin, Dublin 9, Ireland

[#] Current Address: China-Ireland Research Centre for Advanced Materials and Sensor Development, school of Environmental Science and engineering, JiaoTong University, 800 Dongchuan Road, Min Hang, Shanghai, 200240

Correspondence author: kim.lau@dcu.ie, Dermot.Diamond@dcu.ie

Abstract: Carbon dioxide gas concentration determination using infrared gas sensors combined with Bayesian regularizing neural networks is presented in this work. Infrared sensor with a measuring range of 0~5% was used to measure carbon dioxide gas concentration within the range 0~15000 ppm. Neural networks were employed to fulfill the nonlinear output of the sensor. The Bayesian strategy was used to regularize the training of the back propagation neural network with a Levenberg-Marquardt (LM) algorithm. By Bayesian regularization (BR), the design of the network was adaptively achieved according to the complexity of the application. Levenberg-Marquardt algorithm under Bayesian regularization has better generalization capability, and is more stable than the classical method. The results showed that the Bayesian regulating neural network was a powerful tool for dealing with the infrared gas sensor which has a large non-linear measuring range and provide precise determination of carbon dioxide gas concentration. In this example, the optimal architecture of the network was one neuron in the input and output layer and two neurons in the hidden layer. The network model gave a relationship coefficient of 0.9996 between targets and outputs. The prediction recoveries were within 99.9~100.0%.

Keywords: Carbon dioxide; Infrared gas sensors; Neural networks; Bayesian regularization; Levenberg-Marquardt algorithm

1. Introduction

Carbon dioxide is a well known greenhouse gas, which is considered a prime contributor to the climate change mechanism of global warming. Besides the combustion of fuels, landfill gas (LFG) is also an important resource of carbon dioxide. Landfill has always been one of the most attractive disposal routes for municipal solid waste. Over 70% organic solid waste is treated in landfills all over the world. LFG is produced by the breakdown of biodegradable organic materials and comprised of 30-60% methane, 20-50% carbon dioxide and other trace gases depending on the types of waste degrading [1-5]. The majority of the LFG is produced during the working life of a landfill and for a much longer time after a site has been sealed and capped [2].

Environmental pollution arising from the operation and/or management of landfill facilities is now one of the most sensitive social issues in many industrialized countries [3, 4]. Many researchers have paid a great deal of attention to the evaluation of the landfill gas components, specifically carbon dioxide and methane, with the recognition of their potent roles in the global atmospheric environment and/or as an energy resource [1, 5]. Therefore, regular monitoring of landfill gas is required in working and sealed landfills not only to comply with the environmental regulations, but also to judge the need for, or effectiveness of, a landfill control system and to determine the general composition volume over time of air contamination emanating from landfill. Emission rate estimates or modeling may also be used to assess whether it is technically or economically feasible to recover and use the landfill gas for energy production.

Portable equipment is commonly used in landfill to monitor LFG emissions [5, 6]. This monitoring is usually fulfilled monthly by walkover survey. It is time-consuming and labour intensive. Therefore, continuous and automatic monitoring systems are now prevalent among landfills. On the other hand, the great progress in wireless communications in recent years makes it possible to collect long term information on the environment for the monitoring and improvement of these circumstances [7-9]. Various sensors can be used in autonomous monitoring systems for LFG in landfill, such as semiconductor gas sensors, electrochemical sensors and optical sensors [10-14]. Because of a very characteristic infrared absorption wave band, 2360cm^{-1} , corresponding to the C=O stretch, which is seldom interfered by other gases, carbon dioxide molecules can be usually monitored by infrared gas sensors [13]. This technique has an advantage over alternative methods because there is no poisoning mechanism and no need for the presence of oxygen. Additionally, if the infrared source fails, the device fails safely. Since no radiation will reach the receiver, very high gas concentrations will be indicated which warn of a system failure.

Generally, infrared absorption is linearly related to the concentration within a range of analysed gas in ambient air. In the landfill site, however, the concentration of carbon dioxide usually varies in a much larger concentration range of CO₂ gas. In order for the simplification, effectiveness and the cost of a monitoring system, IR sensors with a big enough measuring range, such as 5% or 30%, depending on the purpose of the application, are used in the system in landfills. IR sensors usually linearly respond to the change of the concentration of CO₂ gas within the whole measuring range. However, the response of an IR sensor with a big measuring range to lower concentration gas might deviate from the linearity when it works in a local concentration range of the measuring range. Therefore, the infrared sensors should be calibrated before they can be deployed in the monitoring system in landfill. In this work, we used neural networks to fulfill the calibration of infrared gas sensors, which linear measuring range is 0~5%, for the determination of carbon dioxide within 0~15,000ppm (1.5%).

Neural network

Neural networks are composed of simple elements operating in parallel [15]. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements and can be trained to perform a particular function by adjusting the values of the connections (weights) between elements. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification and classification [16-20]. Neural networks were also well used in environmental applications from LFG production models to calibration of sensors [17].

Commonly, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target [15]. However, one of the problems that occur during neural network training is called over-fitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learnt to generalize for new situations. One method of improving network generalization is to use a network that is just large enough to provide an adequate fit [21]. The larger the network used, the more complex the functions the network can create. If a small enough network is used, it will not have enough power to over-fit the data. Unfortunately, it is difficult to know beforehand how large a network should be for a specific application.

Regularization is usually the first choice for improving generalization among various methods, such as early-stopping or cross validation [21]. This involves modifying the performance function, which is normally chosen to be the sum of squares of the network errors (E_s) on the training set:

$$F(y) = E_s = \frac{1}{N} \sum_{i=1}^N (y_i^S - y_i^{\text{Net}})^2 \quad (1)$$

where y_i^S and y_i^{Net} represent the target and the output of the network for training sample i respectively, and N is the size of the training set.

It is possible to improve generalization if a term that consists of the mean of the sum of squares of the network weights and biases (E_w) is embedded in the performance function [15, 21]:

$$F(y, W) = \alpha E_s + \beta E_w, \quad E_w = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (2)$$

where α, β are performance function parameters, and n is the total number of all weights and biases in the network. The relative size of the performance function parameters dictates the emphasis for training. If $\alpha \gg \beta$, then the training algorithm will produce smaller errors and the over-fitting may still exist. If $\alpha \ll \beta$, training will emphasize weight size reduction at the expense of network errors. Although a smoother network response will be produced, there may be an insufficient fitting for the training set. It is desirable to determine the optimal regularization parameters in an automatic fashion. The main problem with regulation is setting the proper values for the performance function parameters. In this work, a Bayesian framework was employed to automatically adjust the parameters.

Bayesian regulation

The aim of Bayesian regulation is to obtain a neural network with the best performance and the

least parameters so that the problem of over-fitting can be effectively controlled. In order to minimize the regularized performance function $F(y, W)$, the posterior probability of the weights and biases should be maximized on the training data set and a set of α, β . Supposing the prior density for the regularization parameters is uniform, maximizing the posterior density can be achieved by maximizing the likelihood function of the training set. Since the performance function, has the shape of a quadratic in a small area surrounding a minimum point, it can be expanded around the minimum point of the posterior density, W^{MP} , where the gradient is zero. According to Bayesian rules, the optimized regularization parameters at the minimum point can be written as [21, 22]:

$$\alpha^{\text{MP}} = \frac{N - \gamma}{2E_s(W^{\text{MP}})} \quad \text{and} \quad \beta^{\text{MP}} = \frac{\gamma}{2E_w(W^{\text{MP}})} \quad (3)$$

$$\gamma = n - 2\beta^{\text{MP}} \text{tr}(H^{\text{MP}})^{-1} \quad (4)$$

where $H^{\text{MP}} = \alpha \nabla^2 E_s + \beta \nabla^2 E_w$ is the Hessian matrix of the performance function at the minimum point. E_s, E_w are the sum of squares of the network errors and the mean of the sum of squares of the network weights and biases respectively. N is the scale of the training set. γ is called the effective number of parameters, and n is the total number of weights and biases in the network. The parameter γ is a measure of how many parameters in the neural network are effectively used in reducing the performance function. It can range from zero to n .

In the Bayesian framework, the optimization of the regulation parameters requires the computation of Hessian matrix of $F(y, W)$ at the minimum point W^{MP} . In this work, a Levenberg-Marquardt (LM) optimization algorithm [23] is used to estimate the Hessian matrix and locate the minimum point:

$$H = \nabla^2 F(y, W) \approx 2\alpha J^T J + 2\beta I_n \quad (5)$$

where J is the Jacobian matrix whose i -th row equals J_i , The superscript T denotes the matrix transpose, I is the identity matrix. In each step of approximation, the effective number of parameters γ was computed by Eq. 4 with current parameters α, β . After the performance function of the network was minimized by the Levenberg-Marquardt algorithm, a new γ will be obtained and then the new optimized parameters can be estimated by Eq. 3. With each approximation of the performance function parameters the performance function will be changing and the minimum point will keep moving. If traversing the performance surface generally moves toward the next minimum point, then the estimates of the performance function parameters will be more precise. Eventually, the precision will be good enough that the performance function will not significantly change in subsequent iterations. Thus, convergence will be obtained.

In the training procedure, the weights and biases were update according the following equation:

$$W_{k+1} = W_k - (J^T J + \mu I)^{-1} J^T e \quad (6)$$

where e is all errors and I is the identity matrix. μ is an adaptive scalar and k is the training time. When μ is small the algorithm will use the approximate Hessian matrix. When μ is large, this becomes the steepest gradient descent with a small step size. μ decreased by multiplying a coefficient $\mu_{dec} < 0$ after each successful step (reduction in performance function) and increased by

multiplying $1/\mu_{dec}$ only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm. The algorithm stops when μ meets the maximum value predefined.

This work was an attempt to provide an alternative generic technique that could be used to deal with sensor data that are non-linear in nature. We employed an IR sensor that responded non-linearly to carbon dioxide gas to produce a relatively simple data set which could be solved by simple curve fitting methods such as a polynomial curve fitting technique. The plan was to provide a simple and transparent data analysis method as reference to the output produced by ANN. If a more complex situation was used, e.g. with data produced by sensors that did not behave in an ideal manner (unstable sensor responses or drifting sensor responses) or with sensors that produced data that could not be fitted into an easily comprehensible mathematical model, it would be very difficult to validate with confidence the performance of this proposed ANN data processing technique for chemical sensing.

3. Experimental section

3.1 Materials and instrumentation

Infrared (IR) gas sensor (IRgaskiT, Measuring range 0~5%, Edinburgh Instruments Ltd. UK.) was used for the carbon dioxide measurement. Pure dry nitrogen (99.9%) and carbon dioxide (14,900 ppm in nitrogen) (Scott Specialty Gases, UK.) were used to produce gas mixtures of various concentrations controlled by a mass flow controllers (Cole Parmer, Ltd., USA.). An FTIR instrument (Perkin Elmer Ltd. Ireland) was used as the reference.

3.2 Determination of carbon dioxide

The experimental setup is similar to that shown in Figure 1. The experiment was performed at room temperature at 22 °C.

CO₂ concentration measurement was performed with the infrared gas sensor warmed up for 30 minutes. The sensor was connected to a laptop via the USB port for data collection. A range of carbon dioxide samples (shown in Table 2) were obtained in-line by mixing the stock with pure dry nitrogen gas controlled by the mass flow controllers. The gas mixture was passed through the sensor chamber and collected at the outlet to be analyzed by FTIR to certify the concentration of carbon dioxide generated.

The gas samples were delivered at a flow rate of approximate 1 L/min. When the sensor reading had become stable in about five minutes, the whole system was purged with dry pure nitrogen for four minutes to allow the sensor to recover its baseline. The average value of the response plateau was taken as the final output. The experiment was repeated three times. The data was divided into two sets. Sample 1~10 constructed the training set, samples 11~13 as prediction set.

The software program for this work was written in Matlab Version 6.0 (The MathWorks, Inc.). The text file that contained the response captured from the sensor by a PC was fed into the program and the results were written in a text file as output of the program.

4. Results and discussion

4.1. Gas sensor performance

The performance of the gas sensor employed in this work was validated and compared to a conventional bench-top FTIR instrument. A typical response profile of the IR gas sensor to the change of carbon dioxide concentration is presented in the inset of Figure 2 which shows that the sensor responded to the change in gas concentration in the gas chamber reproducibly and rapidly. The sensor response reached equilibrium in about 2 minutes which was actually a measure of the time taken for the sample to equilibrate in the gas chamber. The sensor recovered its baseline in four minutes after the purging with dry pure nitrogen.

The sensor was exposed to a range of predefined CO₂ samples with increasing concentration in three repeats. The same samples were also measured with the FTIR at $\nu_{\max}/\text{cm}^{-1}$ 2360 (corresponding to the C=O bending region) in-line to confirm the actual gas sample concentration. The accuracy of the sample concentration produced by the gas mixer system is shown in Figure 2 by the linear calibration plot obtained from the FTIR instrument. Also presented in Figure 2 is the sensor calibration plot, which is a curved line with an inflexion point at around 7000 ppm. The curved line can be fitted into a polynomial mathematical model as shown in the graph. These results confirmed that the sensor behavior is non-linear within the gas concentration range investigated and that the sensor data are reliable for further analysis with artificial neural network.

4.2. Neural Network approach

Network optimization

A three-layer network was used in this data processing technique. There is one input neuron and one output neuron in the network. The number of hidden neurons was optimized by Bayesian regularization. The transfer functions were selected to be tangent sigmoid for the hidden layer and linear for the output layer. The input and target data were normalized within [-1, 1] to improve the performance of the network. The training goal of error was set as 10^{-6} and maximum generation was 500.

The sensor output data (3 repeats of 13 samples) shown in Figure 2 and Table 2 were processed using ANN with Bayesian regularization that consisted of different number of hidden neurons. Samples #1-10 were used as the training set and samples #11-13 were used as the prediction set.

One of the features of the Bayesian regularization algorithm is the effective number of the network, γ as shown in Eq. 4, which provides a measure of how many network parameters (weights and biases) are being effectively used by the network. If the network has been trained for a sufficient number of generations to ensure convergence, this effective number of parameters γ should reach an approximately constant value, no matter how large the total number of parameters in the network becomes. If γ is very close to the actual number of parameters in the network, then there is a risk that the neural network may not be large enough to properly represent the true relationship between inputs and targets. This can be tested by adding more hidden layer neurons and retrain the new network. If the enlarged network provides a similar γ value, then the former smaller network was actually sufficient. Otherwise, more hidden layer neurons may need to be added until γ reached a steady optimum

values.

The performance of Bayesian regularized neural network with different number of hidden neurons was shown in Table 1. It can be seen that when there was only one neuron in the hidden layer the effective number γ was 3.31, which was close to the total number of parameters of 4 in the network. It can be seen that the sum of squares of the network errors was also very high (88.58×10^{-4}). When the number of hidden neuron was increased to ≥ 2 , the sum of squares of the network errors and the sum of weights and biases significantly decreased. The γ values also reached a relatively steady value with two hidden neurons, which means that the network with 2 hidden neurons was big enough to generalize the involved problem. Therefore, in this work example, the optimal network architecture is one neuron in input layer, two in hidden layer and one in output layer.

In order to investigate the performance of the ANN Levenberg-Marquardt (LM) algorithm without and with Bayesian regularization, all the data was divided into two parts again. Samples # 1, 4, 7, 10-13 was used as the training set with the rest of the samples (# 2, 3, 5, 6, 8, 9) used as the testing set. Training the networks using Levenberg-Marquardt algorithm were performed with two and three neurons in hidden layer. The training iteration for all networks is set at 500, the error goal as 10^{-8} , the maximum μ as 10^{10} and the decreasing coefficient μ_{dec} as 0.1.

The network performances obtained were shown in Figure 3. In general, the networks that used LM algorithm required longer training/testing time; whereas the networks that used BRLM algorithm reached maximum μ at about 400 generations. For the optimal network structure, which has two hidden neurons, smaller training errors were obtained with BR-LM algorithm but the testing errors were slightly smaller for the LM algorithm (Figure 3). It implied that an optimized neural network with LM algorithm has better performance than that with BR-LM algorithm.

Prediction of gas concentration

The final optimized neural network model built after optimized by Bayesian training regularization contained a three-layer forward back-propagation network. There was one neuron in the input layer, two neurons in the hidden layer and one in the output layer. The transfer functions for input and output layers were linear and tangent sigmoid for the hidden layer. The network training error goal was set as 10^{-8} , iteration times as 500. The maximum μ was set as 10^{10} and μ_{dec} as 0.1. Samples #1-10 were used as the training set and samples #11-13 were used as the prediction set. After 480 generations, the training stopped at the training error goal and the neural network model was obtained.

The prediction results and their percentage recoveries obtained were presented in Table 2 and summarized in Figure 4. In the table, it was clearly shown that when parameters were optimized (2 hidden neurons) this ANN technique produced output data (i.e. the predicted CO₂ concentrations) with excellent accuracy where the recovery values were very close to 100%. The relationship coefficient between the target and the output concentrations was 0.999 and the standard deviation was ~27ppm respectively for the training set (Sample 1~10). At 95% confidence, the Fisher test value of the model obtained was 5.4×10^5 .

When the prediction set (Sample 11~13) was presented to the model, very accurate predicted CO₂ concentrations were obtained (Table 2). The recoveries for the three samples are 100.05%, 100.08% and 99.96% respectively. For comparisons, the prediction results by neural networks with 1, 3 and 4 hidden neurons were also listed in Table 2. It can be seen that the prediction errors for the training set

(Sample 1~10) with one hidden neuron were relatively large compared to those with more hidden neurons. With 2 or more hidden neurons, the errors were significantly reduced. Similar results were also obtained with the prediction set (Sample 11~13). These results indicated that the BR-LM algorithm provides very robust model that reduces the errors caused by over-fitting. As a summary, Figure 4 presents a plot of the sensor responses verses the ANN predicted gas concentrations obtained from varying the number of hidden layers and compared to the original input CO₂ concentration. All predicted values were shown to be very close to the original input values within this concentration range suggesting successful application of the proposed ANN algorithm. Inset is a plot of input gas concentration against the predicted concentration calculated from the optimized network. A slope of 0.9988 was obtained demonstrating a very robust and the network which resulted in near-perfect predictions.

In real environment, the precision of the infrared sensor will be affected by the fluctuation of ambient environment such as temperature and humidity etc. These fluctuations will introduce errors to the sensor output and may result in noisy data that would in turn affect the ANN prediction. To simulate this possible problem, a nominal 1% random noise, which was produced randomly by computer, was added to the sensor data to evaluate the robustness of the model. The network was trained by the noisy training data set (Sample 1~10) and used to predict the noisy prediction set (Sample 11~13). The training and prediction were repeated three times where each time new random noises were generated for training. The results are listed in the last column of Table 2. The model obtained from the noisy data sets has a relationship coefficient of 0.9991 and a Fisher test value 2.1×10^5 at 95% confidence. The recoveries for the three noisy prediction samples are 100.00%, 99.89% and 99.62% respectively. This example has demonstrated that there was no significant degradation of the network performance with the addition of 1% random noise which suggests that the neural network model is very robust.

5. Conclusion

Infrared gas sensors can rapidly respond to and measure the concentration change of carbon dioxide gas with recovery in several minutes. However, unlike with a bench-top FTIR system, the response of the IR sensor to the change of concentration of carbon dioxide is nonlinear. A neural network using a Levenberg-Marquardt algorithm with Bayesian regularization was successfully employed to predict unknown CO₂ gas concentrations. Bayesian regularization is an effective way to overcome the problem of over-fitting while training the neural network and this technique provides a generic method for rapid prediction of chemical concentration when non-linear sensors are used.

Acknowledgements:

This research is funded by the Environmental Protection Agency, Ireland under grant code: 2005-AIC-MS-43-M4. Additional funding comes for Science Foundation Ireland under grant code: SFI 03/IN.3/1361. WM Guo and KT Lau express their gratitude to China-Ireland Collaboration Funding Program.

References:

1. L. Lombardi, E. Carnevale, A. Corti, Greenhouse effect reduction and energy recovery from waste landfill, *Energy*, 31 (2006) 3208–3219
2. B. W. Mosher, P. M. Czepiel, R. C. Harriss, J. H. Shorter, C. E. Kolb, J. B. McManus, E. Allwine, B. K. Lamb, Methane emissions at nine landfill sites in the northeastern United States, *Environ. Sci. Tech.*, 33 (1999) , 2088-2094
3. P. Costi, R. Minciardi, M. Robba, M. Rovatti, R. Sacile, An environmentally sustainable decision model for urban solid waste management, *Waste Manag.*, 24 (2004) 277-295
4. S. Squire, M. H. Ramsey, Inter-organisational sampling trials for the uncertainty estimation of landfill gas measurements, *J. Environ. Monitor.*, 3 (2001) 288-294
5. F. V. Almeida, J. R. Guimaraes, W. F. Jardim, Measuring the CO₂ flux at the air/water interface in lakes using flow injection analysis, *J. Environ. Monitor.*, 3 (2001) 317-321
6. K. Spokas, J. Bogner, J. P. Chanton, Morcet M., Aran C., Graff C., Moreau-Le Golvan Y., Hebe I., Methane mass balance at three landfill sites: what is the efficiency of capture by gas collection systems? *Waste Manag.*, 26 (2006) 516-525
7. D. Diamond, Internet-scale sensing, *Anal. Chem.*, 76 (2004) 278A-286A
8. Q. Y. Cai, A. Cammers-Goodwin, C. A. Grimes, A wireless, remote query magnetoelastic CO₂ sensor, *J. Environ. Monitor.*, 2 (2000) 556-560
9. B Kiernan., W. M. Guo, C. Slater, J. Hayes, D. Diamond, Autonomous monitoring of landfill gas migration at borehole wells on landfill sites using wireless technology, *Proceedings of the 10th International Conference on Environmental Science and Technology*, A (2007) 679-685
10. A. B. Kanu, H. H. Hill, M. M. Gribb, R. N. Waters, A small subsurface ion mobility spectrometer sensor for detecting environmental soil-gas contaminants, *J. Environ. Monitor.*, 9 (2007) 51-60
11. G. Hanrahan, D. G. Patil, J Wang, Electrochemical sensors for environmental monitoring: design, development and applications, *J. Environ. Monitor.*, 6 (2004) 657-664
12. J Dickens, M Sepaniak, Modular separation-based fiber-optic sensors for remote in situ monitoring, *J. Environ. Monitor.*, 2 (2000) 11-16
13. J. Mulrooney, J. Clifford, C. Fitzpatrick, E. Lewis, Detection of carbon dioxide emissions from a diesel engine using a mid-infrared optical fiber based sensor, *Sensors and Actuators. A-Phys.*, 136 (2007) 104-110
14. K. Henkel, D. Schmeisser, Back-propagation-based neural network with a two sensor system for monitoring carbon dioxide and relative humidity, *Anal. Bioanal. Chem.*, 374 (2002) 329-337
15. D. J. C. Mackay, A practical Bayesian framework for backpropagation, *Networks: Neural Computation*, 4 (1992) 448-472
16. C. S. Yin, W. M. Guo, S. S. Liu, Y. Shen, Z. X. Pan, L. S. Wang, Simultaneous determination of vitamin B complex using wavelet neural network, *Chinese J. Chem.*, 19 (2001) 836-841
17. D. N. Corrado, D. Fabrizio, D. A. Arnaldo, Pattern recognition in gas sensing: well-stated techniques and advances, *Sensors and Actuators. B: Chem.*, 23 (1995) 111-118
18. C. Y. Li, P. Heinemann, R. Sherry, Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection, *Sensors and Actuators B: Chem.*, 125 (2007) 301-310
19. L. Moreno-Baron, R. Cartas, A. Merkoçi, S. Alegret, M. del Valle, L. Leija, P. R Hernandez., R. Munoz, Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue, *Sensors and Actuators B: Chem.*, 113

(2006) 487-499

20. J. Gallardo, S. Alegret, M. del Valle, A flow-injection electronic tongue based on potentiometric sensors for the determination of nitrate in the presence of chloride, *Sensors and Actuators B: Chem.*, 101 (2004) 72-80
21. J. Lampinen, A. Vehtari, Bayesian approach for neural networks: a review and case studies, *Networks*, 14 (2001) 257-274
22. D. J. C. Mackay, Probable networks and plausible predictions: a review of practical Bayesian methods for supervised neural networks, *Network: Computation In Neural Systems*, (1995) 6 469-505
23. M. T. Hagan, M. Menhaj, Training multilayer networks with the Marquardt algorithm", *IEEE Transactions On Neural Networks*, 5 (1994) 989-993

List of Figures and Tables

Figure 1. Schematic of the experimental setup used for sensor calibration..

Figure 2. Calibration curves for CO₂ samples obtained from IR sensor and FTIR instrument at 2360 cm⁻¹. The FTIR instrument gave a linear plot across the range of gas sample as shown by the solid fitted line. The IR sensor provided a non-linear plot on which a dotted line is added for visual aid purpose only. Inset is the baseline corrected response profile of the IR sensor to different carbon dioxide concentrations in three repeats.

Figure 3. Comparison of the network output using LM and BR_LM algorithm with 2 hidden neurons. (LM: Levenberg-Marquardt algorithm; BR-LM: Levenberg-Marquardt combined with Bayesian regularization algorithm.)

Figure 4. A plot of the artificial neural network predicted CO₂ concentrations using different number of hidden node verses the responses from the IR sensor. Inset is a plot showing the correlation between the input and predicted gas concentration obtained from the optimized network.

Table 1. A summary of the network performance using different number of hidden neurons to show that a maximum effective number (γ) has reached with 2 hidden neurons. Further increasing the network size does not significantly improve the performance.

Table 2. Experimental data and predicted results by neural network. Samples 1-10 were used as the training set and samples 11-13* were the prediction set. '*Noisy data' denotes that the data set has been manipulated by adding a 1% random noise before processed with neural network to test the robustness of the technique.

