Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr 46

XXV Seminarium

ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE 2015

Oddział Gdański PTETiS

METHOD OF SELECTING THE LS-SVM ALGORITHM PARAMETERS IN GAS DETECTION PROCESS

Łukasz LENTKA, Janusz SMULKO

Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics
Department of Metrology and Optoelectronics
tel.:+48 58 348 6095 e-mail: lukasz.lentka@gmail.com

Abstract: In this paper we showed the method of resistive gas sensors data processing. The UV irradiation and temperature modulation was applied to improve gas sensors' selectivity and sensitivity. Noise voltage across the sensor's terminals (proportional to its resistance fluctuations) was recorded to estimate power spectral density. This function was an input data vector for LS-SVM (least squares – support vector machine) algorithm, which predicted a concentration of gas present in sensor's ambient atmosphere. The algorithm creates a non-linear regression model at learning stage. This model can be used to predict gas concentration by recording resistance noise only. We have proposed a fast method of selecting LS-SVM parameters to determine high quality model. The method utilizes a behavior of immune system to determine optimal parameters of the LS-SVM algorithm. High accuracy of the applied method was proved for the recorded experimental data.

Keywords: gas detection, optimal parameters selection, support vector machine, artificial immune system.

1. INTRODUCTION

Gas detection is used in various applications [1-7] (e.g. environmental monitoring systems, food quality recognition, breath analysis). These systems utilize an array of resistive gas sensors and monitor their DC resistances which change according to ambient gas concentration in a non-linear way. Moreover, gas sensing can be improved by modulating temperature or applying UV irradiation which can increase selectivity and sensitivity of the applied gas sensor or reduce their number necessary for proper gas detection [4, 5, 8]. Another possible method of improving gas detection is a fluctuation enhanced sensing (FES) method which measures gas sensors' resistance fluctuations.

Fluctuation enhanced sensing is often more sensitive method than DC resistance measurements only [4, 6, 7]. It can detect a few gases using a single sensor only [9]. However, it requires much more complex measurement setup and more advanced computing [10]. We address these problems in our studies to present nonlinear detection algorithm and a fast way of establishing its parameters.

2. DATA PREPROCESSING

Noise voltage time series are proportional to resistance fluctuation which depend on ambient atmosphere of the investigated gas sensor. The recorded data are used to estimate power spectral density of voltage noise $S_u(f)$ by the equation:

$$S_u(f) = \frac{2}{NT}\sum_{i=1}^{N}\left|X_i(f)\right|^2 \tag{1}$$

where: N – number of the averaged spectra, $X_i(f)$ – Fourier transform of the recorded noise voltage time series, T – observation time of noise time series necessary to calculate a single Fourier transform $X_i(f)$.

The function $S_u(f)$ has to be normalized by the squared sensor bias voltage U_s^2 to be independent from the measurement set-up. Additionally, this product is multiplied by frequency f to expose any disturbance from $1/f$ noise as presented in Fig. 1 [9].

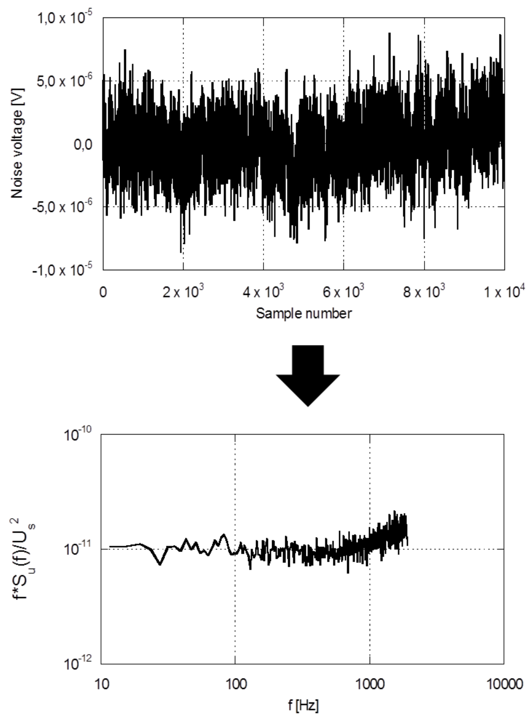

Fig. 1. Conversion and normalization of the recorded voltage noise sequence from time to frequency domain

The spectra product $f \cdot S_u(f)/U_s^2$ can be used as an input data vector for regression algorithm to predict gas concentration. In our work we considered the least square support vector machine algorithm (LS-SVM) because of its limited computation complexity.

3. LS-SVM ALGORITHM

The support vector machine algorithm was proposed by Vapnik [11]. Next, Suykens et al. modified that algorithm to meet least squares criteria and simplify computations by limiting the task to solving a set of linear equations only [12]. Despite this the created model can be non-linear because of the applied non-linear kernel function (e.g. radial basis function – RBF).

Generally the created model for gas concentration prediction can be described by:

$$s = Pc_r + w \tag{2}$$

where: s – set of vectors of the recorded spectra, P – matrix describing the created model, c_r – set of vectors of the measured concentration by a reference method, w – additive white noise component, present during each measurements.

After establishing the model matrix P from a set of the learning data $\{s, c_r\}$ we can predict gas concentration c_p by utilizing newly measured spectra s_n. This stage is a prediction stage and is described by the equation:

$$c_p = P^T s_n + w \tag{3}$$

To obtain high accuracy of gas concentration prediction it is very important to select optimal parameters of the applied LS-SVM algorithm as presented in numerous papers [13, 14]. There are two parameters when the RBF function is applied: γ which determines the compromise between complexity of the model and accuracy of prediction and σ^2 which is the RBF kernel parameter. Moreover, appropriate selection of these parameters guarantees flexibility of the resulting LS-SVM model and protects against model overfitting to the learning data. Additionally, optimal parameters selection affects immunity to additive noise present at each measurement. Authors of the Matlab toolbox LS-SVMLab v1.8 have designed a dedicated function *tunelssvm* which performs optimal selection of this two parameters by utilizing a Nelder-Mead simplex algorithm (which works for all kernels), or a grid search method (restricted to 2-dimensional optimization) or a line search strategy (used with the linear kernel) [15]. Another possible way is to determine these parameters by establishing numerous regression models according to equation (2) and selecting the most accurate model and its parameters. Both methods can secure high accuracy of the established but they are relatively slow. For this reason, we have decided to propose faster optimization algorithm utilizing an Artificial Immune System (AIS) method.

4. ARTIFICIAL IMMUNE SYSTEM

Artificial Immune System is a set of artificial intelligence methods inspired by processes of vertebrate immune system [14, 16]. This methods employs different biological phenomena. In our work we focused on clonal selection algorithm which is commonly applied in various optimization and pattern recognition problems.

To understand the utilized clonal selection process we have to enclose some basic biological principles of that process. The main task of immune system is to protect organism against pathogens which causes diseases. Any factor causing immune reaction is called antigen. Immunity of any organism can be divided into two different types: nonspecific (inborn) and specific (adaptive). The first one cannot be modified or developed, however, the processes occurring in adaptive immunity system are especially interesting due to their continuous adaptation and a feature of preserving information about any recognized antigens. Adaptive immune systems operate using the cells called lymphocytes which circulate around the body and eliminate encountered antigens. The lymphocytes are equipped with antibodies – receptors having ability of recognizing the antigens. Despite of huge number of lymphocytes, they vary to each other and the given antigens can be recognized only by some of them. For this reason the immune system contains a clonal selection mechanism which selects and produces sufficient number of adequate antibodies, necessary to eliminate antigens.

An exemplary algorithm, based on clonal selection principle, is the CLONALG algorithm proposed by de Castro [16]. Figure 2 presents the main idea of that algorithm when applied for optimal selection of the parameters in the LS-SVM algorithm.

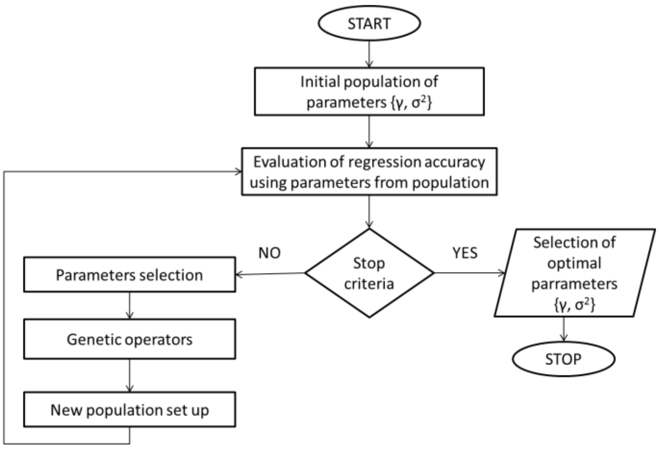

Fig. 2. An algorithm of Clonal selection for establishing optimal parameter in the LS-SVM algorithm

In first step we should establish an initial (random) population of the parameters pairs $\{\gamma, \sigma^2\}$ (e.g. 20 pairs). Next, the LS-SVM models should be created for all the pairs from initial population and the accuracy of these models determined by the root mean squared error (RMSE) should be estimated. Then, the stop criteria is checked – number of performed algorithm iterations (e.g. 30) or achieved detection accuracy (e.g. RMSE < 2 ppm). During the first iteration any of these criterion is not satisfied. In that case only some number of the pairs $\{\gamma, \sigma^2\}$ should be selected (e.g. 5). These selected pairs should guarantee the best regression accuracy. After that they are exposed to genetic operators (mathematical transformations). This operators recalculate each pair with various intensity depending on their matching to the optimal solution of the problem. In this recalculations are also included randomly selected numbers as it is shown in [16], so each evaluation of algorithm can give slightly different solution. In other words, if a pair is more fitted to the optimal solution, it is less modified. Then, a new population of the parameters' pairs is created (e.g. the best 5 pairs from the old set and 15 newly generated by applying genetic operators). The generated population of the parameters $\{\gamma, \sigma^2\}$ is again evaluated in following iterations until reaching criterion of stopping the algorithm (Fig. 2).

5. RESULTS

The presented algorithm was applied to predict concentration of ethanol (C_2H_5OH) and other compounds (C_2H_5OH and NO_2) in gas mixture using a single WO_3 gas sensing layer. Time t_s necessary to select optimal parameters of the LS-SVM regression algorithm in both cases is presented in Table 1. This table includes information about accuracy of the resulting LS-SVM model expressed as the RMSE value σ.

Table 1. Comparison of parameters' selection time t_s and accuracy σ of the resulting LS-SVM model obtained using different methods

	tunelssvm		AIS	
	t_s [min]	σ [ppm]	t_s [min]	σ [ppm]
Ethanol	5.23	1.71	1.33	1.75
Gas mixture	20.58	6.07	7.87	7.19

All calculations were performed using the same exemplary computer. The *tunelssvm* function from the Matlab toolbox gives slightly better accuracy of ethanol prediction but time of parameters selection was much longer. Thus, we obtained about four times faster selection of optimal parameters for the LS-SVM regression algorithm for the same computer as previously reported. Figure 3 presents a graph of C_2H_5OH concentration prediction when the sensor was in ambient atmosphere of ethanol at different concentrations. The reference (real) value of gas concentration was set using flowmeters.

The presented data were obtained for the prototype gas sensing layer of WO_3 nanoparticles, using condition as described in the literature [17].

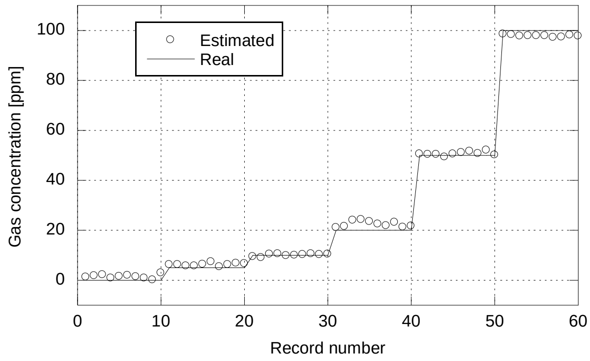

Fig. 3. Results of C_2H_5OH concentration prediction in WO_3 gas sensing layer using LS-SVM regression algorithm

For prediction of C_2H_5OH concentration and NO_2 concentration in gas mixture, the *tunelssvm* function gave slightly better accuracy but time necessary for selecting optimal parameters was much longer (see Table 1).Thus, the proposed method assured about three times faster parameters selection. Figure 4 presents results of concentration prediction of both components in the investigated gas mixture. The reference values of gases concentrations were set using flowmeters.

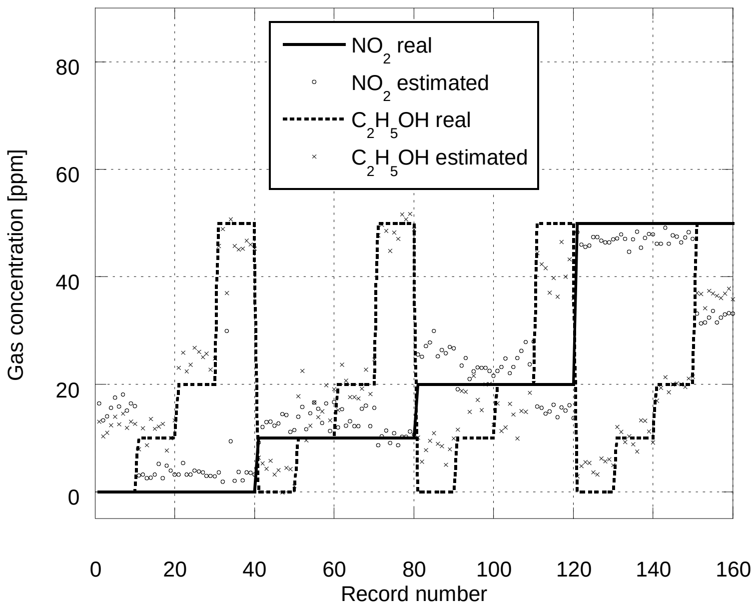

Fig. 4. Results of gas mixture compounds concentration prediction using WO_3 gas sensing layer and the LS-SVM regression algorithm

The RMSE values from Table 1 were established for testing data only, for learning data those values were almost the same, so they aren't included in the table.

Table 2 presents complexities (numbers of support vectors) of established models using different methods. The optimal parameters selection method doesn't effect complexity of the model. For gas mixture the resulting model is more complex than for only one gas (ethanol).

Table 2. The number of support vectors in different cases

	tunelssvm	AIS
Ethanol	240	240
Gas mixture	640	640

6. CONCLUSIONS

We showed that fluctuation enhanced sensing together with the LS-SVM algorithm results in high accuracy of gas concentration prediction. A clonal selection algorithm was used to speed up the process of determining optimal parameters for the LS-SVM regression algorithm when the RBF kernel function was applied. We have experimentally confirmed that the applied methods (fluctuation enhanced sensing and the applied regression algorithm) can determine more than one chemical compounds present in the gas mixture when a single resistive sensor (WO_3 prototype gas sensing layer [17]) was used.

There are numerous applications for the presented methods. Especially in portable gas detection systems where a rough estimation of concentration is needed and abilities of intense computing are limited. The applied algorithm should help to decrease costs of the hardware by simplifying necessary computations.

ACKNOWLEDGEMENT

This work was supported by the National Science Center, Poland, the grant decision: DEC-2012/06/M/ST7/00444 "Detection of gases by means of nano-technological resistance sensors".

7. BIBLIOGRAPHY

1. Zakrzewska K.: Mixed oxides as gas sensors, Thin Solid Films, vol. 391, 2001, pp. 229-238.
2. Osowski S., Siwek K., Grzywacz T., Brudzewski K.: Differential electronic nose in on-line dynamic measurements, Metrology and Measurement Systems, vol. 21, 2014, pp. 649-662.
3. Kalinowski P., Woźniak Ł., Strzelczyk A., Jasiński P., Jasiński G.: Efficiency of linear and non-linear classifiers for gas identification from electrocatalytic gas sensor, Metrology and Measurement Systems, vol. 20, 2013, pp. 501-512.
4. Trawka M., Smulko J., Hasse L., Ionescu R., Annanouch F., Llobet E., Granqvist C., Kish L.: Fluctuation enhanced gas sensing using UV irradiated Au-nanoparticle-decorated WO_3-nanowire films, Proceedings of the 8th International Conference on Sensing Technology, Liverpool, September 2014, pp. 282-286.
5. Heszler P., Ionescu R., Llobet E., Reyes L., Smulko J., Kish L., Granqvist C.: On the selectivity of nanostructured semiconductor gas sensors, Physica Status Solidi B, vol. 244, 2007, pp. 4331-4335.
6. Kwan C., Schmera G., Smulko J., Kish L., Heszler P., Granqvist C.: Advanced agent identification with fluctuation enhanced sensing, Sensor Journal, IEEE, vol. 8, 2008, pp. 706-713.
7. Kotarski M., Smulko J.: Hazardous gases detection by fluctuation-enhanced gas sensing, Fluctuation and Noise Letters, vol. 9, 2010, pp. 359-371.
8. Lee A., Reedy B.: Temperature modulation in semiconductor gas sensing, Sensors and Actuators B, vol. 60, 1999, pp. 35-42.
9. Kish L., Vajtai R., Granqvist C.: Extracting information from noise spectra of chemical sensors: single sensor electronic noses and tongues, Sensors and Actuators B, vol. 71, 2000, pp. 55-59.
10. Kotarski M., Smulko J.: Noise measurement set-ups for fluctuations-enhanced gas sensing, Metrology and Measurement Systems, vol. 16, 2009, pp. 457-464.
11. Vapnik V.: The nature of statistical learning theory, Springer Science & Business Media, New York, 2000.
12. Van Gestel T., De Brabanter J., De Moor B., Vandewalle J., Suykens J.: Least squares vector machines, World Scientific, Singapore, 2002.
13. Nguyen M., De la Torre F.: Optimal feature selection for suport vector machines, Pattern Recognition, vol. 43, 2010, pp. 584-591.
14. Aydin I., Karakose M., Akin E.: A multi-objective artificial immune algorithm for parameter optimization in suport vector machine, Applied Soft Computing, vol. 11, 2011, pp. 120-129.
15. Pelckmans K., Suykens J., Van Gestel T., De Brabanter J., Lukas L., Hamers B., De Moor B.: LS-SVMlab: a matlab/c toolbox for least squares support vector machines, 2002.
16. De Castro L., Von Zuben F.: Learning and optimization using the clonal selection principle, Evolutionary Computation, IEEE Transactions on, vol. 6, 2002, pp. 239-251.
17. Lentka Ł., Smulko J., Ionescu R., Granqvist C., Kish L.: Determination of gas mixture components using fluctuation enhanced sensing and the LS-SVM regression algorithm, Metrology and Measurement Systems, vol. 22, 2015, pp. 341-350.

SPOSÓB DOBORU PARAMETRÓW ALGORYTMU LS-SVM W PROCESIE DETEKCJI GAZÓW

W artykule pokazano metodę przetwarzania danych z rezystancyjnych czujników gazów, stosowaną do wykrywania gazów. W celu zwiększenia czułości i selektywności czujników zastosowano modulację temperaturową oraz oświetlenie diodą LED UV aby zebrać więcej danych. Szumy napięciowe rejestrowane na zaciskach czujnika (proporcjonalne do fluktuacji jego rezystancji) zostały wykorzystane do wyznaczenia gęstości widmowej mocy. Ta funkcja stanowiła wektor danych wejściowych dla algorytmu maszyny wektorów nośnych według kryterium najmniejszych kwadratów (LS-SVM), umożliwiając określenie stężenia gazu występującego w atmosferze otaczającej czujnik. Nieliniowy charakter algorytmu pozwala na tworzenie w fazie uczenia modelu na podstawie danych uzyskanych z pomiarów za pomocą metody odniesienia. Pokazano szybki sposób doboru optymalnych parametrów algorytmu LS-SVM, gwarantujących skuteczność szacowania stężenia gazu. W badaniach wykorzystano metodę symulującą działanie systemu odpornościowego. Analiza danych eksperymentalnych potwierdziła skuteczność prezentowanej metody.

Słowa kluczowe: detekcja gazów, optymalny dobór parametrów, maszyna wektorów nośnych, sztuczny system immunologiczny.