Available online at www.sciencedirect.com
ScienceDirect
Sensors and Actuators B 122 (2007) 227–235
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

Probabilistic support vector machines for multi-class alcohol identification

F.J. Acevedo [a,*], S. Maldonado [a], E. Domínguez [b], A. Narváez [b], F. López [a]

[a] *Department of Signal Theory and Communications, University of Alcalá, 28871 Madrid, Spain*
[b] *Department of Analytical Chemistry, University of Alcala, 28871 Madrid, Spain*

Received 20 January 2006; received in revised form 12 May 2006; accepted 22 May 2006
Available online 11 July 2006

Abstract

In this work we address the use of support vector machines in multi-category problems. In our case, the objective is to classify eight different kinds of alcohols with just one SnO_2 sensor using thermomodulation. The use of support vector machines in the field of sensors signals recognition is beginning to be used due to the ability to generalize in a binary classification problem with a small number of training samples. However, when a multi-class problem is presented, the outputs of the support vector machines are uncalibrated and should not be used to determine the category. In this work a step forward is added to the output of the binary classifiers to choose the category with a maximal a posteriori probability. Obtained results show that the ability of generalization provided by support vector machines improves the results obtained with other learning methods used in the electronic nose field and their use in multi-class problems can be addressed with the method proposed. To reduce the high dimensionality of the data we have benchmarked several feature extraction methods with probabilistic support vector machines.

Keywords: Electronic nose; Support vector machines; Feature extraction; Thermomodulation

1. Introduction

Electronic noses are defined as an array of sensors and a pattern recognition (PARC) system [1,2]. Over the past years these systems have been applied to many different applications [2], from coffee quality analysis [3] to automotive exhaust constituent measurements [4]. Especially important are those systems based on SnO_2 sensors due to their advantages of low-manufacturing cost and high sensitivity. However, disadvantages of these sensors come from the lack of stability and drift [6]. This is a major problem because it makes the response non-stationary and the performance of the system in a real situation is poorer than in a controlled environment.

Most of the works have researched on the dynamic response of the signal in order to extract more information and to minimize the drift effect. Basically, there are two approaches to this transient response: varying the concentration of the analyte [7] or varying the working temperature [8] also known as thermomodulation. Each approach then results in another signal. The second approach has been applied to SnO_2 sensors and has resulted in more stability. Moreover, it has been demonstrated [9] that every sensor is more sensitive to a specific gas at a particular temperature. It is therefore possible to obtain dynamic responses in time which then act as a "fingerprint" for a specific vapour [10]. With this idea in mind and with the help of a PARC system it is possible to identify the presence of different substances. However, although the stability increases with this technique there is always a number of erroneous measurements, also known as outliers. In order to improve the classification, the PARC system selected has to be robust in the presence of these outliers.

Support vector machines (SVM) is a classification technique based on statistical learning theory (SLT) [11,12]. The use of SVM in the field of sensors signals classification for two categories is beginning to be used [13–15] due to the ability to generalize a classification problem with a small number of training samples and so, to avoid the mentioned effects due to the outlier samples. Since SVM solve two-class problems,

* Corresponding author. Tel.: +34 91 885 6725; fax: +34 91 885 66 99.
E-mail address: javier.acevedo@uah.es (F.J. Acevedo).

0925-4005/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2006.05.033

the extension to the multi-class case is usually done with different binary SVM that separates one substance from the rest, as is done in Ref. [16]. However, the output of a SVM is an uncalibrated value and should be not compared directly. Instead, Platt [17] proposed to estimate the a posteriori probability given the output of a SVM which in turn makes it possible to compare between different binary classifiers.

In this work, our proposal is to classify several kinds of alcohols with just one SnO_2 sensor using probabilistic support vector machines such as the PARC system. Alcohols to be classified are based on pure alcohols (ethanol, propanol and methanol), aromatic alcohols (veratryl, isoamyl, amyl) and complex structures presented in colognes or liquors, eight different substances in all. Alcohol identification has been reported in some previous works [18–20] but although they used thermomodulation it was necessary to use several sensors. What is more, they identified less substances than we propose to. We have compared the results obtained with SVM with those obtained with some PARC schemes used in the electronic nose such as Fuzzy Artmaps [21,22] or Neural Networks [23].

One of the aspects gaining in importance in the electronic nose field is feature extraction [24,25]. These techniques try to reduce the dimension of the input vectors to be recognized by the learning machine employed, concentrating all the discriminating information on a few features only. We have also tested a set of these methods to study the behaviour of SVM in each case.

Although the described work classifies alcohols, the procedure here exposed can be applied to different applications in the field of classification of substances by their odour.

2. Experimental

2.1. Signal acquisition

The sensor stage is composed of a signal conditioning circuit, a power stage and a unique SnO_2 from Figaro Sensor (Taguchi 2620). This sensor incorporates a heater element. In our case it is fed through the power stage with a sinusoidal heater voltage in order to achieve thermomodulation. A frequency of 50 mHz was selected for this modulating signal. This frequency must be slow enough to allow the heater element to change the temperature following the heater voltage. If a much higher frequency is applied, the heater will not be able to change the temperature and thermomodulation is not achieved. In the presence of alcohol the sensor changes its resistance giving a dynamic response due to thermomodulation. This modulated response is sent to a microcontroller that converts the signal from analog to digital with a resolution of 10 bits.

For the training phase we have to collect a representative group of signals of every kind of substance, we would like to identify—alcohols in this case. These signals are all acquired under the same thermomodulation signal in different environmental conditions of temperature and humidity. The pure alcohols and the aromatic ones were tested without having diluted them in water, since we wanted to classify the alcohols rather than estimate the concentration. The experimental setup of the sampling vapours was done placing the liquid into a closed chamber where the sensor is located. This chamber is not kept at constant temperature because, as it was mentioned, we would like to test the system under different environmental conditions. For the training stage, the liquids were placed without having diluted them. The purpose of this first phase is to test the proposed classification method without interference effects from other substances. The training of the system is always done with these substances without having diluted them.

For the test stage the procedure followed to sampling the vapours was quite similar to the one described for the training stage, but instead of using substances without having diluted them, we also tested the substances diluted at three different concentrations because it is interesting to evaluate the detection limit. We also tested some mixes from two different substances.

Every cycle of the thermomoduling signal is considered a realization in the time domain of the alcohol substance representation and is represented as $v_{s_i}^k[n]$, where s_i $i = 1, 2, \ldots, 8$ is one of the alcohol substances under study, and k represents the kth cycle of the thermomodulated signal. So, for a N cycles signal we have N realizations of one of the substances under study, with T samples of each one. This can be seen in Fig. 1a for four cycles of amyl alcohol. The fact of acquiring the signals under different environmental conditions results in that the signals taken with the same substance are not very similar, as is shown in Fig. 1b–i. where we can distinguish different profiles for each of the substances under study. Each of the cycles obtained is converted to a T-dimension vector which in turn is the input of the PARC system.

2.2. Autonomous real-time system

The method exposed in this work has been implemented in an autonomous real-time system based on a custom board designed by our investigation group, incorporating a Fujitsu 16-bit microcontroller, the signal acquisition part and an external host. This system can work in two modes: training and testing. In the training mode, the system sends the signal, captured in the sensor stage described above, to an external workstation that finds the optimal values for the multi-category problem. The software that is executed in the workstation is based on a written-in-house version of the SVMLIB [26]. The mode's objective is to produce a configuration file with the PARC coefficients that will be loaded into the autonomous system ready for the test mode. Once the training phase has ended the external host is no longer needed.

In the testing mode, the system has been updated with the necessary information to know which of the signal's features need to be extracted. Furthermore, the system has been fed the coefficients of the support vector machines and the a posteriori probability coefficients to identify the alcohol. The result is sent to a display that informs the final user.

It is important to note that the available memory in this kind of systems is small and this plays an important role in the implementation of the system. As the memory is limited we should think about possible solutions during the design process and test some feature extraction algorithms that allow the information stored in the memory to be small.

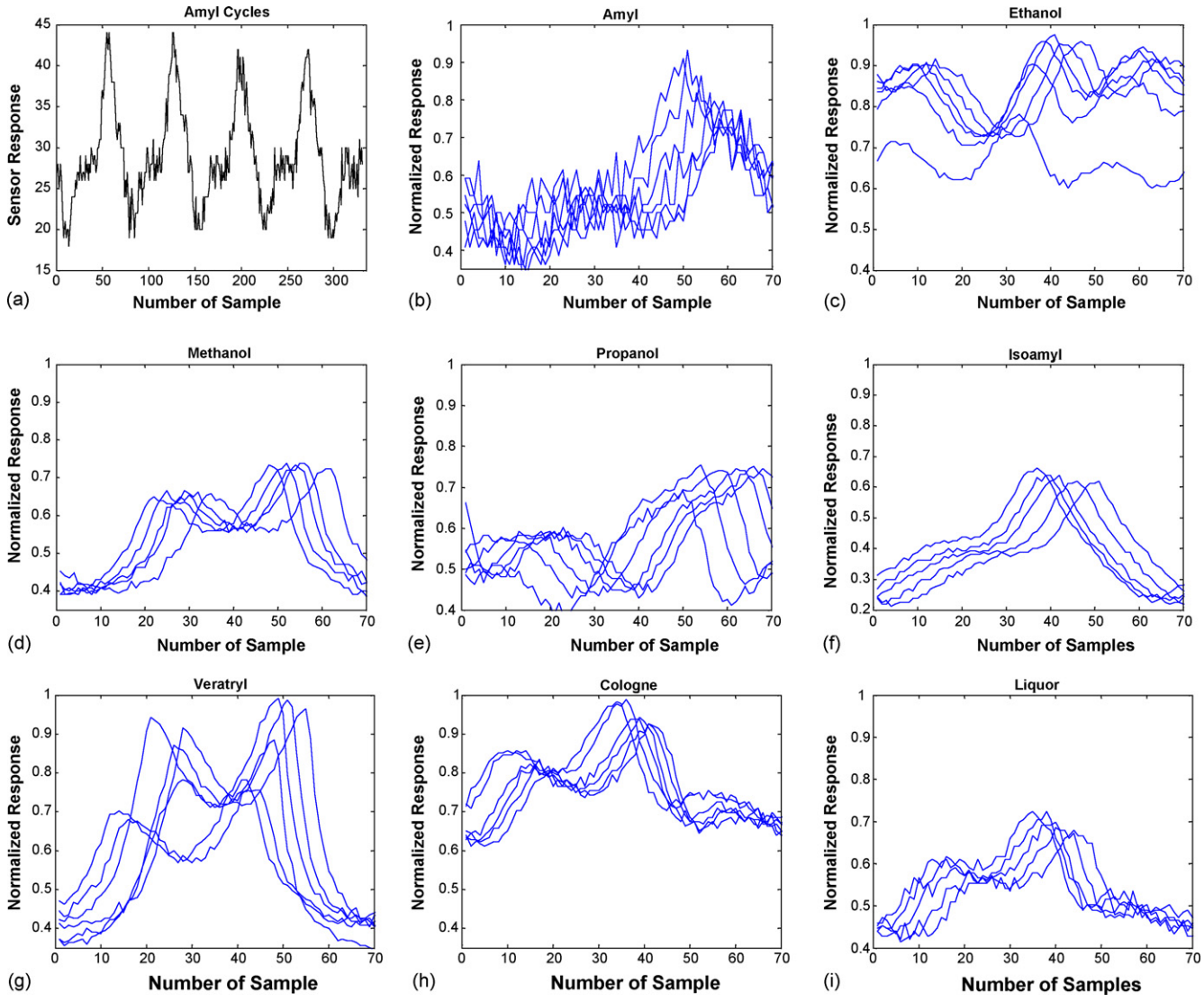

Fig. 1. (a) Themomodulated signal of the amyl substance. (b–i) Different realizations with changes in the environment for the following substances: (b) amyl, (c) ethanol, (d) methanol, (e) propanol, (f) isoamyl, (g) veratryl, (h) cologne and (i) liquor.

3. Support vector machines for multi-class problems

3.1. Support vector machines for two-class problems

SVM theory comes from the statistical learning theory and is based on the structural risk minimization (SRM) rather that on the empirical risk minimization (ERM) as is done in the neural networks theory. Let us suppose we have a set of l training patterns $\{\boldsymbol{x}_i, y_i\}$ where $i = 1, 2, \ldots, l$. $y_i \in \{+1, -1\}$ and $\boldsymbol{x}_i \in \mathbb{R}^d$. In our case, vectors $\boldsymbol{x}_i$ are the sensor responses during one cycle or vectors composed by the features extracted from those responses, whereas y_i are the labels associated to each pattern indicating one of the classes. Thus, a two-class problem basically consists of finding the optimal hyperplane that separates the samples labelled -1 from those labelled $+1$, with a given margin between one set and the other. Such a hyperplane is found when the margin is maximal. Instead of solving this optimization problem, in Ref. [12] we can find that is easier to solve the dual problem:

$$\begin{aligned} &\min_{\alpha} L_D = \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_i \alpha_i \\ &\text{subject to :} \\ &0 \le \alpha_i \le C \\ &\sum_i \alpha_i y_i = 0 \end{aligned} \tag{1}$$

Being α the Lagrange multipliers of the primal optimization problem of obtaining maximum margin, C is a regularization parameter and $K(\boldsymbol{s_i}, \boldsymbol{x})$ is a kernel function that computes the inner product in a higher dimension space. This kernel function should satisfy Mercer's Theorem to ensure that the optimization problem exposed is convex [12]. In this work, we have used a radial function basis (RBF) kernel for the binary classifiers,

defined as:

$$K(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{e}^{-\gamma\|\boldsymbol{x}-\boldsymbol{y}\|^2} \tag{2}$$

Note, that from the set of training vectors only for a reduced group, the associated coefficients α_i are different to zero. These training vectors with a non-null coefficient value are called support vectors. The output of a binary classifier is then calculated as:

$$f_x = \sum_{i=1}^{N_s} \alpha_i y_i K(\boldsymbol{s}_i, \boldsymbol{x}) + b \tag{3}$$

where N_s is the number of support vectors found as a result of the optimization problem, $\boldsymbol{s}_i$ the support vectors and b is a threshold parameter updated in the training phase. The classification rule depends on the sign of f_x. It must be noted that this output is an uncalibrated value and the outputs of two different classifiers should not be compared.

3.2. Multi-class strategies for SVM

Most of the classification problems in the sensors signal analysis are multi-category problems, as the alcohol substances classification problem exposed here. It seems very interesting to extend the good performance of the binary SVM to multi-class problems.

Weston and Watkins [27] proposed a formulation of the optimization problem considering all classes at once. Although this strategy is very elegant from a mathematical viewpoint, its implementation does not yield so good results as in the binary case.

Other popular strategies in the pattern recognition field are to consider the problem as a collection of binary classification problems. In the one-against-all strategy k classifiers are constructed. The kth classifier constructs a hyperplane between the class k and the rest of the classes. In the one-against-one strategy $(k(k-1)/2)$ hyperplanes are defined, where each hyperplane separates each class from the other. Although this second strategy sometimes improves the results obtained with the one-against-all strategy, it should be noted that the number of SVM increases very much, even in the case where the number of categories is small.

In our case, the PARC had to be implemented in the autonomous real-time system described, with eight different classes to discriminate. It should be noted that every support vector machine requires a quantity of memory calculated as the number of support vectors multiplied by the dimension of the problem. So, the one-against-one strategy consumes more memory. This fact makes the one-against-all strategy the most adequate one for our case.

3.3. Probabilistic SVM for multi-class classification

Let us suppose we address our multi-category classification problem by constructing a set of binary classifiers under the premise one-against-all. In that case, every classifier discriminates one alcohol substance from the rest. Once we have trained the eight classifiers, an input is given to the classifiers to test which class this input belongs to. However, the ethanol classifier and the methanol one, for example, could both give a positive output. In such a case, what classifier should be taken into account? The optimal choice from a statistical viewpoint is to select the classifier whose output has maximal a posteriori probability. However, if the binary classifiers are based on SVM, as mentioned before, the output f_x for a given input $\boldsymbol{x}$ is an uncalibrated value, and so it is not possible to compare the outputs of the binary classifiers. Consequently, under a multi-class problem based on binary SVM classifiers, we need to add a stage that transforms the output f_x to a posteriori probability $P(y = 1/f_x)$.

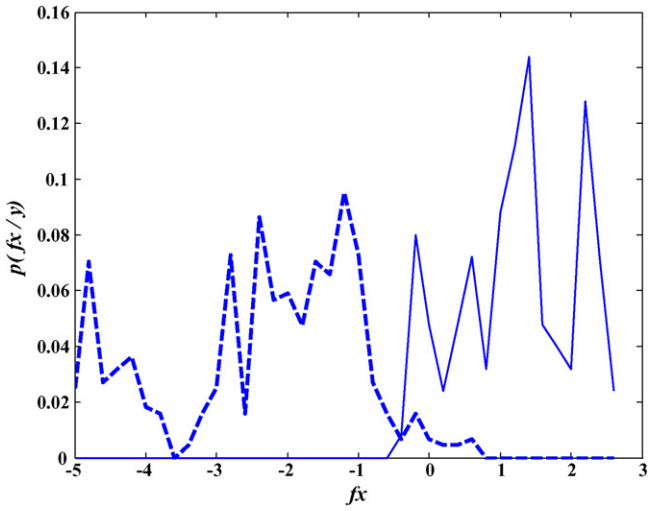

Fig. 2. Ethanol classification histograms for $p(f_x/y = -1)$ as a dashed line and $p(f_x/y = 1)$ as a continuous line.

To make this calculus of the a posteriori probability possible, we begin by estimating the probability density functions (pdf) of $p(f_x/y = 1)$ and $p(f_x/y = -1)$ from the training set. In the multi-class problem described in this paper we select the binary classifier that separates ethanol from the rest of substances. This classifier is based on a SVM trained with a RBF kernel. In Fig. 2 the pdf $p(f_x/y = 1)$ is shown as a continuous line whereas the pdf $p(f_x/y = -1)$ is represented as a dashed line. Hastie and Tibsharini [28] proposed to approximate these two pdf to both Gaussian distributions, but as we can see in Fig. 2 this assumption is not very accurate for our alcohol classification problem.

Once these pdf have been estimated the Bayes rule is applied to obtain the a posteriori probability, giving for the ethanol classifier the graph shown in Fig. 3:

$$P\left(y = \frac{1}{f_x}\right) = \frac{p(f_x/y = 1)P(y = 1)}{\sum_{i=1,-1} p(f_x/y = i)P(y = i)} \tag{4}$$

Platt [17] proposed to use a parametric model to adjust this a posteriori probability by means of a sigmoidal function. In the previous figure we can appreciated that, given the set of points, a sigmoid function is a reasonable election. So, instead of directly estimating the a posteriori probability we will estimate

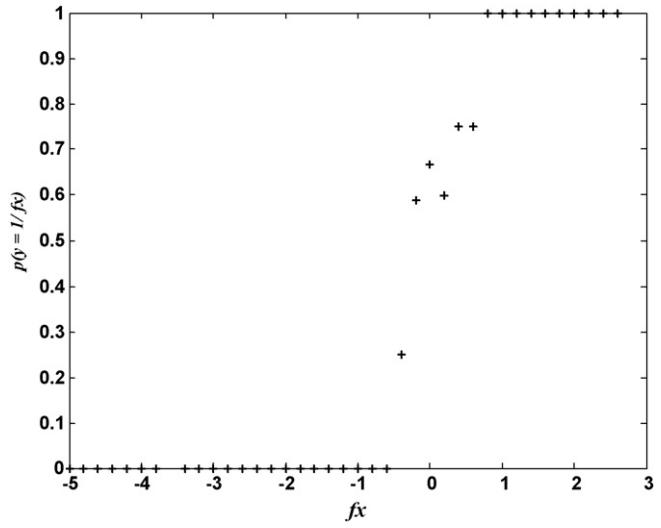

Fig. 3. $P(y=1/f_x)$ for the ethanol SVM classifier.

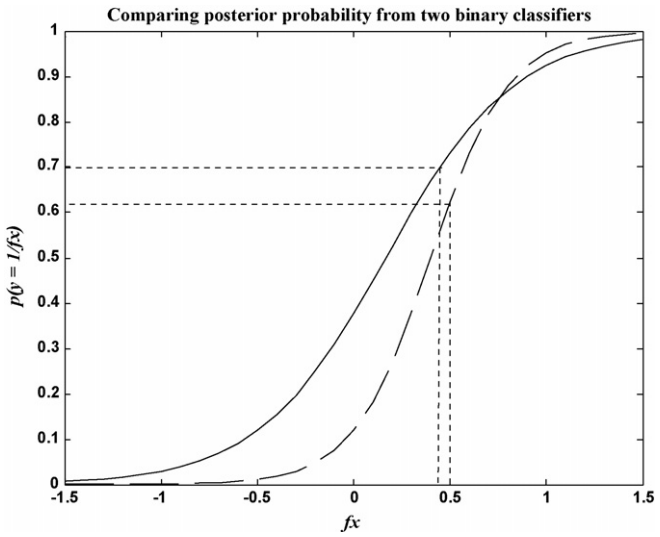

Fig. 4. Sigmoides for two classifiers and the problem of selecting the one with a greater f_x value.

the parameters A and B of a sigmoid:

$$P\left(y=\frac{1}{f_x}\right)=\frac{1}{1+\exp(Af_x+B)} \tag{5}$$

These parameters can be found using any regularization method. In our case, we followed Platt's criteria to minimize the following function based on the likelihood ratio:

$$\min F(A,B)=-\sum_{i=1}^{l} t_i \log\left(\frac{1}{1+\exp(Af_{x_i}+B)}\right) + (1-t_i)\log\left(1-\frac{1}{1+\exp(Af_{x_i}+B)}\right) \tag{6}$$

$$t_i=\begin{cases}\dfrac{N_+ +1}{N_+ +2}, & \text{if } y_i=1\\ \dfrac{1}{N_- +2}, & \text{if } y_i=-1\end{cases}$$

where l is the number of training patterns, N_+ the number of positive training patterns and N_- is the negative training points. The advantage of using this criterion to adjust the sigmoid parameters is that the Hessian $H(A,B)=\nabla^2 F(A,B)$ is positive definitive and so the problem can be solved using Newton methods without any risks of finding local minima.

Once this theory has been exposed it is possible to understand the mistake which results from directly comparing the outputs of several binary classifiers. Let us suppose that we have two sigmoid functions obtained as exposed in Fig. 4. The sigmoid drawn as a continuous line is associated to the binary classifier that detects the presence of methanol, whereas the one drawn as a dashed line is associated to the propanol classifier. For this example, let us suppose that, given an input pattern $\boldsymbol{x}$, a negative value is obtained for the rest of the classifiers, while values of 0.5 and 0.45 are obtained for the propanol and methanol classifiers. If the greatest output is selected, we would pick the propanol category. However, in this situation the a posteriori probability of the methanol category is greater, being this selection the one that we should choose. This example also shows that if the values of parameters A and B are quite similar, the direct output of the SVM criterion would work fine, although this criterion makes no sense from a probability point of view.

4. Results and discussion

4.1. Sigmoid estimation

As the classification problem is a multi-class one, the sigmoid parameters described in the previous section must be estimated before benchmarking the SVM. The simplest procedure to do this is to train every binary classifier with a training set of samples to obtain the alpha multipliers and thus, the decision function. Once the hyperplane has been calculated, we can use the training set itself to minimize the likelihood function described in Eq. (6).

However, the result would be biased by the training set and affected by the fact that a lot of training samples would become support vectors and so, when testing them the classification error would be null. In this work we have followed a jack-knife approximation, that is, for the same binary set problem we separated a group of samples and used them only for test purposes. The SVM is constructed with the rest of the samples and the separated set provides us with some values to be able to train the sigmoid. To obtain more values to train the sigmoid, we now take out other different set of samples for test purposes and we then repeat the process. It is well known that the jack-knife approach gives an unbiased estimator but introduces a great computational complexity when the number of samples taken as test samples is small, being the extreme case the leave-one-out approach.

For the example under study, the training set was composed of 849 samples. The values of the A parameters were estimated using a RBF kernel with the gamma parameter equal to 0.8 and by dividing the training set five times. These values are reflected in Table 1, whereas the B values are not reflected because all of them were zero.

Table 1
Sigmoid parameters estimation

Substance	*A* value
Ethanol	−2.8
Metanol	−4.2
Propanol	−5.4
Liquor	−4.2
Veratryl	−11.6
Isoamyl	−7.9
Amyl	−5.8
Cologne	−78.6

4.2. Comparing Multi-Class SVM to other PARC methods

Once sigmoid parameters have been estimated and the binary classifiers are trained, we can use Multi-Class SVM as PARC system in our multi-category problem. To test the behaviour of the proposed classifier, we compared it to other PARC systems widely used in the field of the electronic nose. As the problem under study is not easily separable, we compared the system with neural networks and fuzzy ARTMAP. To compare them, the accuracy parameter is used, defined as:

$$\rho = \frac{1}{K}\sum_{i=1}^{K} P\left(\frac{D_i}{H_i}\right)$$

where K is the number of classes, in our case eight and $P(D_i/H_i)$ means the probability of correct classification for the class i. Table 2 shows the accuracy of each PARC obtained with an independent set of test samples composed of 105 samples obtained from substances without having diluted them.

SVM were constructed using a RBF kernel and the sigmoid values indicated in Table 1. The Neural Network employed was a multi-layer perceptron (MLP) with 20 neurons in the hidden layer and 10,000 epochs for the training phase. The number of neurons in the hidden layer was selected by trying with different numbers and picking the best result. Fuzzy Artmap has a good behaviour compared to the simplicity of the algorithm and requires fewer operations than the others in the testing phase. However, in this case, it is possible to appreciate the improvement of the results when using SVM with a RBF kernel. Table 3 shows the confusion matrix using a SVM with an RBF kernel.

Table 2
Accuracy obtained with different PARC systems without preprocessing

Alcohol type	Learning machine		
	SVM RBF	Neural network	Fuzzy ArtMap
Ethanol	0.975	0.85	0.95
Metanol	0.975	0.33	0.46
Propanol	0.975	0	1
Liquor	0.975	1	0.66
Veratryl	0.975	1	1
Isoamyl	0.975	0.5	0.7
Amyl	0.866	1	1
Cologne	0.800	0	0
Mean accuracy	0.9395	0.58542	0.72292

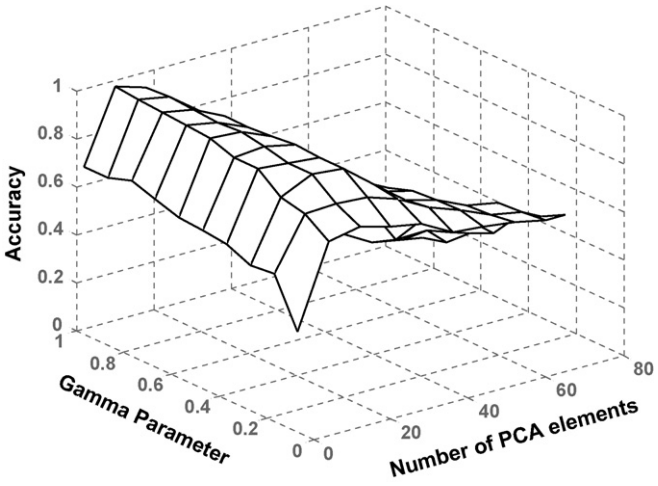

Fig. 5. Accuracy against gamma and number of PCA features.

4.3. Feature extraction benchmarking

Feature extraction is a very interesting previous stage that transforms the input patterns, in order to reduce the problem's dimensionality. This stage is of major importance because working with a smaller dimension means a drastic reduction in the number of operations and this could be critical in a real-time system. For this propose, four methods were tested coupling them to SVM. A sweep of the γ parameter of the RBF kernel was done to view the influence of this parameter on the accuracy obtained.

The first feature extraction method proposed is the principal component analysis (PCA). Fig. 5 shows the accuracy against the number of features extracted and the gamma parameter of the RBF kernel. We realise that this method reaches a high grade of accuracy when the number of features extracted is between 10 and 20, decreasing when more features are considered.

The second method tested is the linear discriminant analysis (LDA). This method is also widely used and employs data transformation to maximize the relation between inter-dispersion and intra-dispersion. In Fig. 6 we can appreciate the accuracy against

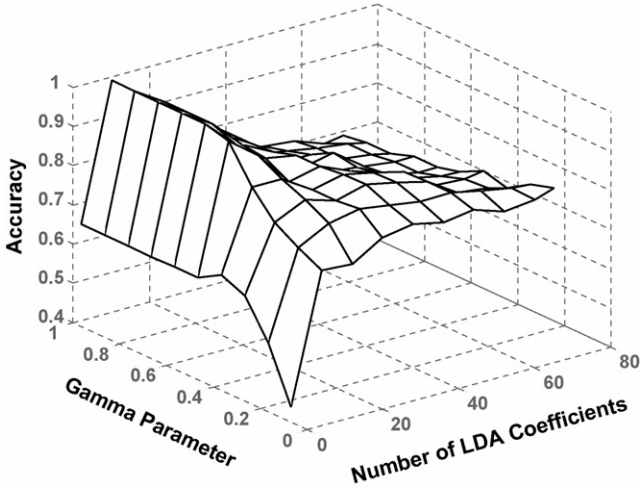

Fig. 6. Accuracy against gamma and number of LDA features.

Table 3
Confusion matrix for the RBF SVM case

	Ethanol	Methanol	Propanol	Liquor	Veratryl	Isoamyl	Amyl	Cologne
Ethanol	15	0	0	0	0	0	0	0
Methanol	0	15	0	0	0	0	0	0
Propanol	0	0	15	0	0	0	0	0
Liquor	0	0	0	10	0	0	0	0
Veratryl	0	0	0	0	15	0	0	0
Isoamyl	0	0	0	0	0	10	0	0
Amyl	0	0	0	0	2	0	13	0
Cologne	0	0	0	1	0	1	0	8

gamma. Both methods, PCA and LDA, use linear transformations based on the statistics of the training set and give satisfactory results when coupled to a SVM with the appropriated gamma value. However, as we have said before, we are working with an autonomous system and both methods require storing the transformation matrix in the memory, but as the problem dimensionality is reduced, we come upon advantages when storing support vectors in the memory.

In the other two feature extraction methods the transformation is fixed. These methods are the discrete cosine transform and the discrete wavelet transform, and are very popular in the field of data compression. In this work, the criterion followed to indicate what coefficients were most important was the mean energy of the transformed coefficients of the training set. Results are shown in Figs. 7 and 8 and we can see that more features are necessary to reach the high level of accuracy obtained with the PCA and LDA methods. However, the main advantage is that it is not necessary to store any transformation matrix in the memory, only the filters that are fixed (Table 3).

4.4. Results for different concentrations

It has been shown the good behaviour of the proposed method compared to other PARC system when the substances are not diluted. However, it is interesting to test the system when the substances are diluted. Thus, it is possible to check the importance of the concentration and to calculate the detection limit.

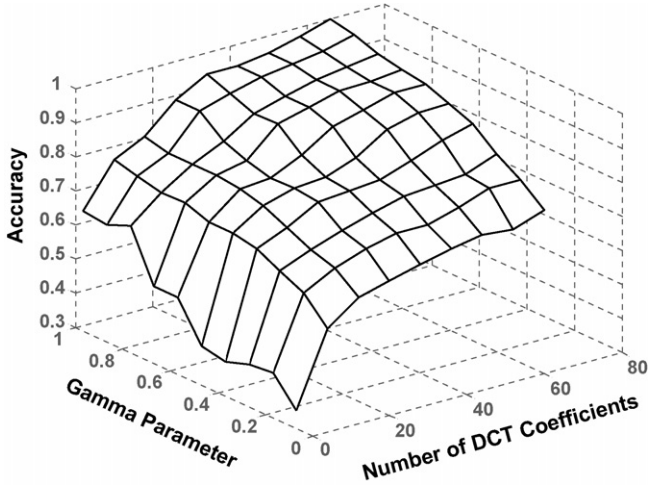

Fig. 7. Accuracy against gamma and number of DCT coefficients.

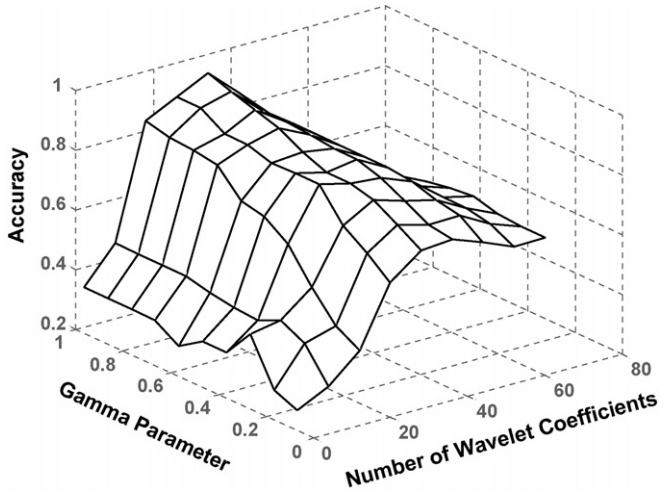

Fig. 8. Accuracy against gamma and number of DWT coefficients.

For this proposal, the substances were diluted in water, except for the amyl case, at three different concentrations. For each substance and for each concentration 50 samples were taken as it has been described in the experimental setup. It is important to note that the training coefficients have not change, because it was decided not to re-train with the new substances since we want to know the detection for not seen substances. In Table 4 it is shown the percentage of success for each concentration.

It can be appreciated that the method has a good behaviour when high concentrations are used. However, when the substances are diluted in with a 75% of water the fingerprints obtained are quite similar between them and the system is not able to classify them with high percentage of success. Anyway, this lack of accuracy is higher when we tested these substances with other PARC methods. If the application we are interested.

Table 4
Percentage of success for each substance at different concentrations

	75%	50%	25%
Ethanol	0.94	0.82	0.66
Methanol	0.96	0.8	0.62
Propanol	0.96	0.84	0.42
Liquor	0.84	0.72	0.5
Veratryl	0.88	0.72	0.6
Isoamyl	0.92	0.86	0.76
Amyl	0.82	0.62	0.46
Cologne	0.74	0.62	0.4

5. Conclusions

The main contribution of the paper is the use of support vector machines for multi-class problems based on binary classifiers and how to deal with the output in order to allow a voting scheme. It should be mentioned that in many cases good results can be achieved by directly taking the SVM outputs and proposing a voting scheme. As mentioned before, this can be done when the sigmoid training for all the classes indicates that the risk associated with doing this task directly is low. However, from a mathematical viewpoint, the a posteriori probability stage should be added in all cases.

On the other hand, we have tested some popular feature extraction methods coupling them to support vector machines. Results indicate that for the application under study, the dimension of the problem can be reduced with higher accuracy levels.

Future work will be to test other kernels and to formulate a coherent feature extraction system with the support vector machines theory. The system has to be improved to work at low concentrations, being interesting to test new sensor technologies.

Acknowledgments

This work has been supported by the Comunidad de Madrid projects GR/MAT/0916/2004 and CAM-UAH 2005/031.

References

[1] R. Gutierrez-Osuna, Pattern analysis for machine olfaction: a review, IEEE Sens. J. 2 (3) (2002) 189–202.
[2] E.L. Hines, E. Llobet, J.W. Gardner, Electronic noses: a review of signal processing techniques, IEE Proc. Circuit Devices Syst. 146 (1999) 297–310.
[3] J.W. Gardner, P.N. Barlett, Electronic Noses: Principles and Applications, Oxford University Press, 1999.
[4] J.W. Gardner, H.V. Shurmer, T.T. Tan, Application of an electronic nose to the discrimination of coffees, Sens. Actuators B, Chem. 6 (1992) 71–75.
[6] K. Ihokura, J. Watsin, Stannic Oxide Gas Sensors, Principles and Applications, CRC Press, Boca Raton, FL, 1994.
[7] X. Vilanova, E. Llobet, R. Alcubilla, J.E. Sueiras, X. Correig, Analysis of the conductance transient of thick-film tin dioxide gas sensors under successive gas injection steps, Meas. Sci Technol. 8 (1997) 1133–1138.
[8] A. Heilig, N. Bârsan, U. Weimar, W. Göpel, Selectivity enhancement of SnO_2 gas sensors: simultaneous monitoring of resistances and temperatures, Sens. Actuators B, Chem. 58 (1999) 302–309.
[9] S.W. Wlodek, K. Colbow, F. Consadori, Signal-shape analysis of a thermally cycled tin-oxide gas sensor, Sens. Actuators B, Chem. 3 (1991) 63–68.
[10] S. Al-Khalifa, S. Maldonado, J.W. Gardner, Identification of CO and NO_2 using a thermally resistive microsensor and support vector machine, IEE Proc.-Sci. Meas. Technol. 150 (6) (2003) 11–14.
[11] V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.
[12] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Mining Knowledge Discov. 2 (1998) 121–167.
[13] S. Maldonado-Bascón, Feature reduction using support vector machines for binary gas detection, in: Proceedings of the Seventh International Work-Conference on Artificial and Natural Neural Networks, Menorca, 2003.
[14] J. Acevedo, S. Maldonado, S. Al-Khalifa, J.W. Gardner, P. Gil, Design of a programmable, portable and low-cost electronic nose, in: Proceedings 11th International Symposium on Olfaction and Electronic Noses, Barcelona, 2005.
[15] M. Pardo, G. Sberveglieri, Classification of electronic nose data with support vector machines, Sens. Actuators B, 107 Chem. 107 (2005) 730–737.
[16] C. Distante, N. Ancona, P. Siciliano, Support vector machines for olfactory signals recognition, Sens. Actuators B, 107 Chem. 88 (2003) 30–39.
[17] J. Platt, Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, in: Advances in Large Margin Classifiers, MIT Press, 2000.
[18] R. Ionescu, A. Hoel, C.G. Granqvist, E. Llobet, P. Heszler, Ethanol and H_2S gas detection in air and in reducing and oxiding ambience: application of pattern recognition to analyse the output from temperatue-modulated nanoparticulate WO_3 gas sensors, Sens. Actuators B, Chem. 104 (2005) 124–131.
[19] A. Fort, N. Machetti, S. Rocchi, M.B. Serrano, L. Tondi, N. Ulivieri, V. Vignoli, G. Sberveglieri, Tin oxide gas sensing: comparison among different measurement techniques for gas mixture classification, IEE Trans. Instrum. Meas. 52 (2003) 921–926.
[20] A. Gramm, A. Schütze, High performance solvent vapor identification with a two sensor array using temperature cycling and pattern classification, Sens. Actuators B, Chem. 95 (2003) 58–65.
[21] E. Llobet, J. Brezmes, R. Ionescu, X. Vilanova, S. Al-Khalifa, J.W. Gardner, N. Bârsan, X. Correig, Wavelet transform and fuzzy ARTMAP-based pattern recognition for fast gas identification using a micro-hotplate gas sensors, Sens. Actuators B, Chem. 83 (2002) 238–244.
[22] E. Llobet, E.L. Hines, J.W. Gardner, Fuzzy ARTMAP-based electronic nose data analysis, Sens. Actuators B, Chem. 61 (1999) 183–190.
[23] M. Aleixandre, I. Sayago, M.C. Horrillo, M.J. Fernández, L. Arés, M. García, J.P. Santos, J. Gutiérrez, Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas, Sens. Actuators B, Chem. 103 (2004) 122–128.
[24] C. Distante, M. Leo, P. Siciliano, K.C. Persaud, On the study of feature extraction methods for an electronic nose, Sens. Actuators B, Chem. 87 (2002) 274–288.
[25] A. Leone, C. Distante, N. Ancona, K.C. Persaud, E. Stella, P. Siciliano, A powerful method for feature extraction and compression of electronic nose responses, Sens. Actuators B, Chem. 105 (2005) 378–392.
[26] C. Chang, C. Lin, LIBSVM: A Library For Support Vector Machines, 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[27] J. Weston, C. Watkins, Multi-Class Support Vector Machines, Technical Report CSD-TR-98-04, Department of Computer Science, University of London, 1998.
[28] T. Hastie, R. Tibshirani, Classification by Pairwise Coupling. Technical Report, Standford University and University of Toronto, 1996.

Biographies

D. Francisco Javier Acevedo received the MS degree in electronic engineering from University of Alcalá in 1998. For several years he has been involved in many projects in the private industry. Since 2004, he is a member of the Signal Theory and Communications Division of the University de Alcalá. His research interest range from processing signals from sensors to pattern recognition and real time systems.

Dr. Saturnino Maldonado received the telecommunication engineering degree from the Universidad Politécnica de Madrid (UPM), Madrid (Spain) in 1996 and the PhD degree. Since 1997, he has been a member of the Signal Theory and Communications Division of the University de Alcalá, becoming an associate professor in 2001 at the Department of Signal Theory and Communications. His research interests are image processing and pattern recognition.

Dr. Arántzazu Narváez obtained her European PhD degree in 1999. Presently, she is holding a National research contract (Ramón y Cajal program) at the Bioelectrochemistry group in the University of Alcalá. She has published 25 international papers. She has also worked at different laboratories including the Hebrew University of Jerusalem (Prof. I. Willner), the Ecole Normale Superieure (Paris, Prof. C. Amatore), the University of Lund (Sweden, Prof. L. Gorton),

and the University of Dortmund (Germany, Prof. C.M. Niemeyer). Her research interest range from molecular biology to bioanalytical devices focussing in DNA sensors.

Prof. Elena Domínguez studied Pharmacy at the University Complutense of Madrid. She obtained her doctorate in 1982 in the University of Alcalá. During almost 5 years, she worked at Lund University in Sweden and returned to Spain in 1991. Since 2002, she is professor of analytical chemistry at the University of Alcalá. For the last 15 years, she has leaded the Bioelectrochemistry group focussing on surface modification, the characterisation of catalytic and affinity events, bioelectrochemistry and transducing chemistry for the development of reliable biosensing systems in the area of diagnostics, food and environmental analysis. Different kinds of bioreceptors (enzymes, antibodies, nucleic acids and nuclear receptors) have been explored with mainly electrochemical means of transduction and using optical and microgravimetric techniques for sensor characterization, been this work crystallised in more than 75 international publications.

Dr. Francisco López Ferreras received the MS degree in telecommunication engineering from the Universidad Politécnica of Madrid in 1970 and the PhD degree in 1980. Currently, he is Dean at the Engeneering Politechnique Department in the University of Alcalá. His research interests lie in the area of signal processing, where he has been author or co-author in 23 journals, 16 book chapters, 88 conference contributions and 10 books.