Available online at www.sciencedirect.com
SCIENCE DIRECT
Sensors and Actuators B 116 (2006) 174–182
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

A combined pattern separability and two-tiered classification approach for identification of binary mixtures of VOCs

Robi Polikar [a,*], Kauser Jahan [b], Bryan Healy [a]

[a] *Signal Processing and Pattern Recognition Laboratory, Electrical and Computer Engineering, Rowan University, 136 Rowan Hall, 201 Mullica Hill Road, Glassboro, NJ 08028, USA*
[b] *Civil & Environmental Engineering, Rowan University, 136 Rowan Hall, 201 Mullica Hill Road, Glassboro, NJ 08028, USA*

Received 10 July 2005; received in revised form 17 October 2006; accepted 18 November 2006
Available online 18 April 2006

Abstract

Several classification techniques have been developed with varying degrees of success for automated identification of VOCs, however, the problem becomes considerably more challenging when more than one VOC is present. The reason is two-fold: first, the response of the sensors to certain VOCs may be too strong and mask the response of the sensors to other VOCs in the environment; and second the responses of the sensors to VOCs may not have enough separability information if the specificity of the sensors is not adequate. We propose the following procedures for these two issues in identification of binary mixtures of VOCs: a nonlinear cluster transformation technique or nonparametric discriminant analysis to increase pattern separability, followed by a two-tier classification to aid in identification of dominant and secondary VOCs separately. Results demonstrate the feasibility of the combined approach.

Keywords: Binary VOC mixture identification; Electronic nose; Neural networks; Pattern separability

1. Introduction

Volatile organic compounds (VOC) can be found in a variety of settings, including industrial (such as wastewater) as well as residential (such as drinking water supplies, hospitals) sites. Most VOCs find their way into the environment through human causes such as industrial pollution or fuel spills. These compounds can cause disastrous effects on the environment through premature degradation of the surrounding area and health hazards to people living around the contaminated areas. The need for an accurate, cost-effective and objective system for detection and identification of VOCs is therefore undisputed.

Various laboratory based methods exist for the examination of water or air samples and detection of these compounds, but they are often expensive and not located near the source of the pollution, causing difficulties in continuous testing. Systems employing electronic noses (Enose) have recently become popular for on-field detection and classification, saving both time and money [1]. The data generated by the Enose can then be analyzed by a pattern recognition system for automated identification of the VOC present in the environment. In fact, several such techniques have been proposed over the last decade for identification of (single) VOCs, each with varying degrees of success [2–5]. However, the problem becomes considerably more challenging, when the VOCs appear in a mixture, and the individual components of this mixture need to be identified [6–8]. This difficulty arises from primarily two factors: first, in many cases the sensor(s) may have a very similar response to two very different compounds, a direct result of inadequate specificity of the sensor. Second, the sensor response to one of the components in the mixture may be so strong that the response to the other components may be completely masked. In this study, we consider binary mixtures of such VOCs, where the VOC generating the stronger response is referred to as the dominant VOC, whereas the other is called the secondary VOC.

We present a composite approach to the above mentioned two issues. We propose a classification system that first applies a preprocessing algorithm, nonlinear cluster transformation or nonparametric discriminant analysis, in order to increase the

* Corresponding author. Tel.: +1 8562565372; fax: +1 8562565241.
E-mail address: polikar@rowan.edu (R. Polikar).

0925-4005/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2005.11.079

separability of the data to aid in classification. We then use a two-level classification system: first the separability algorithm is applied to raw data followed by a neural network classifier to determine the dominant VOC only. Then, a second level of neural network – specifically tuned to the dominant VOC chosen during the first step – is used to determine the secondary VOC in the mixture.

To date, on a database that includes 24 binary combinations of five dominant and seven secondary VOCs, we have obtained performance figures mid-80% to mid-90% on identification of both dominant and secondary VOCs, while the classifiers were unable to converge if either of the pattern separability or two-tier classification was not employed.

2. Experimental

2.1. Experimental setup

Piezoelectric acoustic wave sensors comprise a versatile class of chemical sensors for the detection of VOCs. Addition or subtraction of molecular material from the surface or bulk of an acoustic wave sensor results in a change in its resonant frequency. The frequency change, Δf, caused by a deposited mass Δm, can be described by the Sauerbrey equation [9,10]. For quartz crystal microbalances (QCMs), this relationship is given by

$$\Delta f = -2.3 \times 10^6 \frac{\Delta m}{A} f_0 \quad (1)$$

where f_0 is the fundamental resonant frequency of the bare crystal, and A is the sensing surface area. For sensing applications, a sensitive polymer film is cast on the surface of the QCM. This film can bind the molecules of the VOC of interest, altering the resonant frequency of the device in proportion to the added mass. The QCM-based chemical sensor system typically consists of an array of several crystals, each coated with a different polymer film. The response pattern of such an array then serves as the signature for a given VOC. This array design is aimed at improving identification, which is hampered by the limited selectivity and sensitivity of individual films.

The electronic nose system used to generate the data analyzed in this study is an array of six 9 MHz quartz crystal microbalances (QCM). Cr/Au contacts were first deposited onto the quartz, to serve as electrodes, by means of a resistive heating evaporator, and the sensors were subsequently coated with different polymer films, chosen to maximize the specificity of the sensor array for the specific VOCs of interest.

An array of six crystals, coated with the following polymers, was used to detect and identify twenty-four binary mixtures of VOCs. The polymers were APZ: apiezon L, PIB: poly(isobutelene), DEGA: poly[di(ethylene glycol) adipate], SG: solgel, OV275: poly[bis(cyanoallyl)polysiloxane], PDS: poly(dimethylsiloxane) and PDPP: poly(diphenoxyphosphazene). The films were dilute solutions of these polymers, typically 20 μL of 0.3–3% (w/w), spinning at 2000–5000 rpm. The sensors were then dried at 65 °C for 24 h. The coated QCMs were subsequently mounted in a sealed test fixture, which could house up to six sensors.

Fig. 1 depicts a schematic of the experimental setup. The vapor generation system consisted of a carrier gas, typically dry nitrogen flowing at 200 sccm, a gas stream module, and a pair of three-way switchable valves. The three-way valves were used to maintain constant flow of the reference gas and the VOC. The final output was a constant gas flow rate with periodic exposure of the array to known levels of VOC. Flow was maintained by using calibrated mass flow controllers (Tylan general FC-280 AV) and conventional gas bubblers containing the analytes. The bubblers were composed of two connected compartments. The gas carrier bubbled through the solution in the first compartment, supplying the vapor, whereas the second analyte-containing compartment served as a headspace equilibrator. The vapor was further diluted with nitrogen to obtain the desired concentration levels. The sensors were exposed to the vapor stream by means of computer controlled three-way valves and a MKS multi-gas controller model 147B that electronically controlled the mass flow controllers. Polyethylene and Teflon®

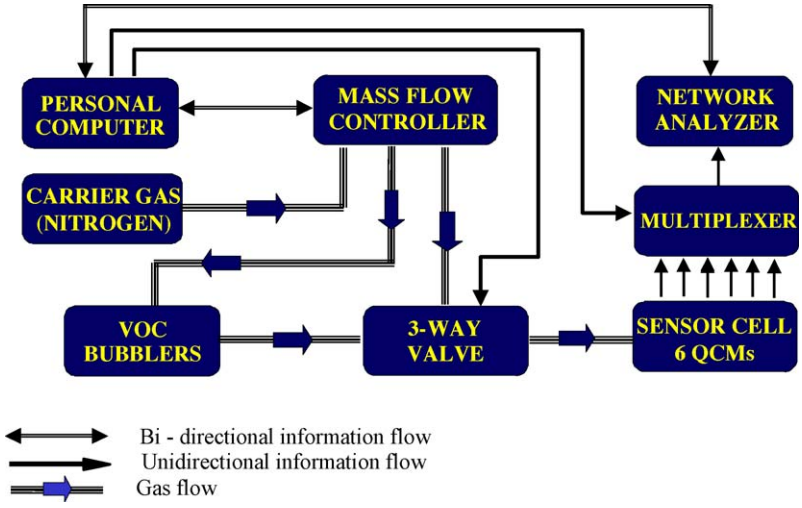

Fig. 1. Experimental setup for VOC detection.

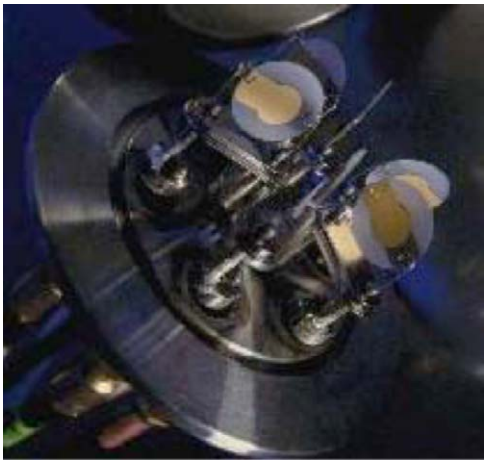

Fig. 2. QCMs housed in a sensor cell.

tubings together with stainless steel or brass valves were used, but only Teflon® and stainless steel were exposed to the analytes. All experiments were performed at ambient temperature.

Repeated measurements indicated reproducibility of the collected data with insignificant variations of 2–4%. The variability, due to small temperature fluctuations, was within experimental error. The frequency response was monitored using a HP8753C network analyzer, interfaced to an IEEE 488 card installed in a PC, running HP8516A resonator-measurement software. Real time data were displayed and saved. The data were then analyzed to obtain frequency shifts (relative to the baseline) versus VOC concentration. Typical noise levels (standard deviations of the baseline) for the QCMs were around 0.01 Hz. Fig. 1 depicts the overall schematic of this setup, whereas Fig. 2 shows the QCMs installed in the sensor cell.

2.2. Data collection and handling

The database generated in this study consisted of 24 binary mixtures of the following nine VOCs: acetonitrile (ACN), methyl ethyl ketone (MEK), ethanol (ET), 1,1,1-trichloroethane (TCA), trichloroethylene (TCE), hexane (HX), octane (OC), toluene (TL) and xylene (XL). Table 1 lists the binary mixtures obtained from these VOCs.

Among the VOCs used in this study, ET, TCE, OC, and in particular, TL and XL, constitute the dominant VOCs. Each column represents mixtures of one of the five dominant VOCs, with one of other secondary VOCs. Sensors were exposed to these mixtures at all combinations of 150, 300, 500 and 700 parts per million (ppm), giving 16 combinations of concentrations for each of the 24 mixtures listed above (that is, 150 and 150, 150 and 300, 150 and 500, 150 and 700, 300 and 150, . . ., 700 and 500, 700 and 700 ppm). Twenty-four mixtures of 16 different combinations of concentrations generated the 384-pattern database used in this study.

As described next, even a careful selection of polymers did not provide a well-separated feature space for the identification of VOCs.

Table 1
Binary mixtures of VOCs analyzed in this study

Octane	Xylene	Toluene	TCE	Ethanol
OC & ACN	XL & ACN	TL & ACN	TCE & TCA	ET & ACN
OC & ET	XL & ET	TL & ET	TCE & MEK	ET & MEK
OC & MEK	XL & MEK	TL & MEK	TCE & TL	ET & HX
OC & TL	XL & HX	TL & HX	TCE & ET	ET & TCA
OC & TCA	XL & TCA	TL & TCA	TCE & HX	

3. Increasing pattern separability

3.1. Pattern separability problem for identification of mixtures of VOCs

In general, classification algorithms work well when the classes are well separated in the feature space, which requires that the underlying data distributions have adequately large intercluster distances between patterns of different classes, and small intracluster distances between patterns of the same class. For most real-world problems, however, this is rarely the case as overlapping class patterns in the feature space is usually the norm rather then the exception.

For identification of binary mixtures of VOCs, there are two problems that need to be addressed: first, the patterns generated by the sensors overlap considerably in the feature space with small intercluster and large intracluster distances. Second, the existence of a dominant VOC in the mixture masks the response to the secondary VOC. The following paragraphs illustrate these problems.

Table 2 summarizes the actual frequency shifts (in Hz) and the corresponding normalized values for two mixtures: XL/MEK mixture and TL/HX mixture. All responses were normalized with respect to amplitude through square root of the sum of squares of sensor responses (prior to any processing) to remove concentration dependent information and ensure signature pattern based identification. This is because, in real-world environments, the concentrations of the individual analytes are unknown, and hence the identification cannot be based on the concentration information. Yet, this normalization makes the identification exceedingly difficult: despite the significantly different actual frequency shifts between the two mixtures, the normalized responses are extremely close to each other.

Fig. 3 shows the bar graphs (patterns) corresponding XL & MEK and TL& HX mixtures. The vertical axis is the normalized relative frequency change in response to VOC molecules

Table 2
Original and normalized frequency shift responses of two mixtures

	XL&MEK original (Hz)	TL&HX original (Hz)	XL&MEK normalized	TL&HX normalized
APZ	290	95	0.30	0.29
PIB	793	264	0.81	0.81
DEGA	154	54	0.16	0.17
SG	172	60	0.18	0.18
OV	138	51	0.14	0.16
PDPP	411	139	0.42	0.42

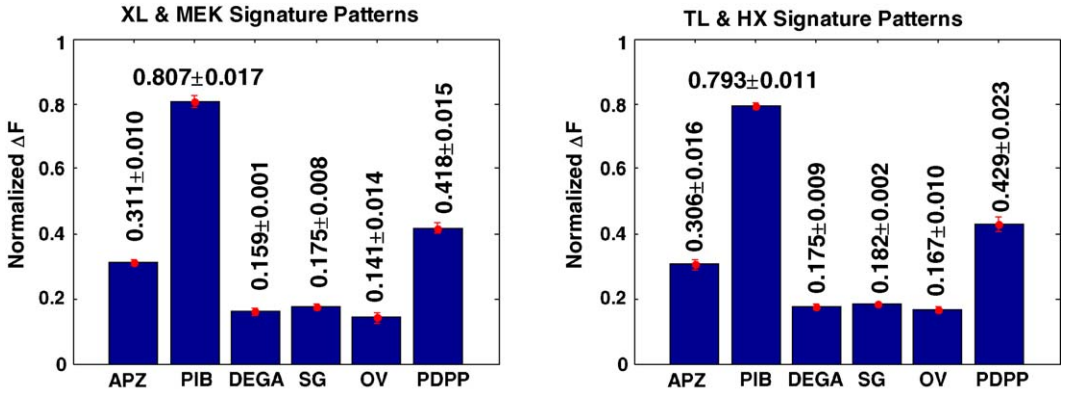

Fig. 3. Sample QCM responses to different mixtures.

deposited on the sensor surface, averaged over all normalized responses, along with the one-standard-deviation intervals of the average values. Each bar represents the average normalized response of a sensor coated with the specified polymer. The six normalized numerical values corresponding to the sensor responses for each mixture then constituted the six-dimensional feature vectors used in classification. The similarity of response patterns for different mixtures is quite striking, presenting a considerable challenge to the classification algorithm. Also note that the one-standard-deviation intervals are extremely small, indicating very little variation in the normalized responses.

Fig. 4 shows typical patterns obtained from four different mixtures of xylene. These patterns illustrate the second difficulty associated with this particular application. Even if one of the two components of the mixture is known, identifying the second VOC is still a significant challenge. This is due to the dominance of responses to xylene in comparison to others. When a dominant VOC is present in the mixture, the responses of sensors to other VOCs become partially, or sometimes completely, masked by the response to the dominant VOC, resulting in highly packed (and possibly overlapping) clusters in the pattern space. This observation is further confirmed by Fig. 5 which shows the average response patterns of xylene, toluene, methanol and ethanol individually. We note that xylene response is very similar to those of xylene mixtures in Fig. 4, and furthermore is very similar to the toluene response alone. This shows that xylene

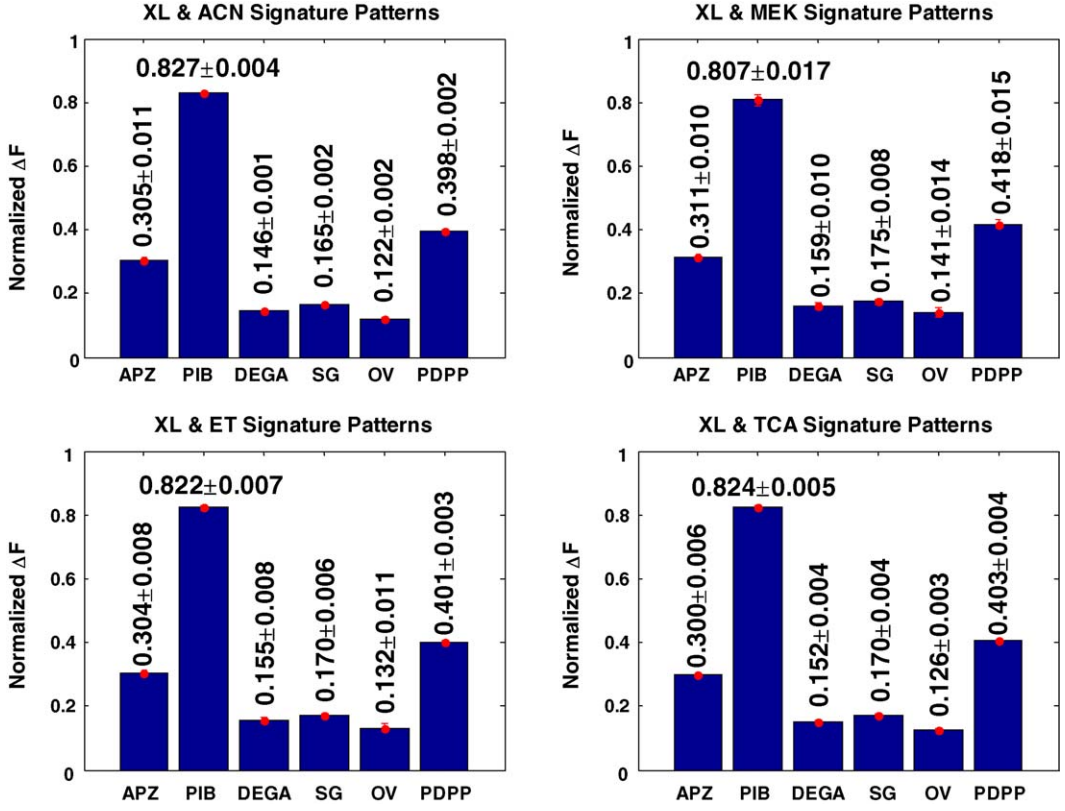

Fig. 4. QCM responses: (a) XL&ACN, (b) XL&MEK, (c) XL&ET and (d) XL&TCA.

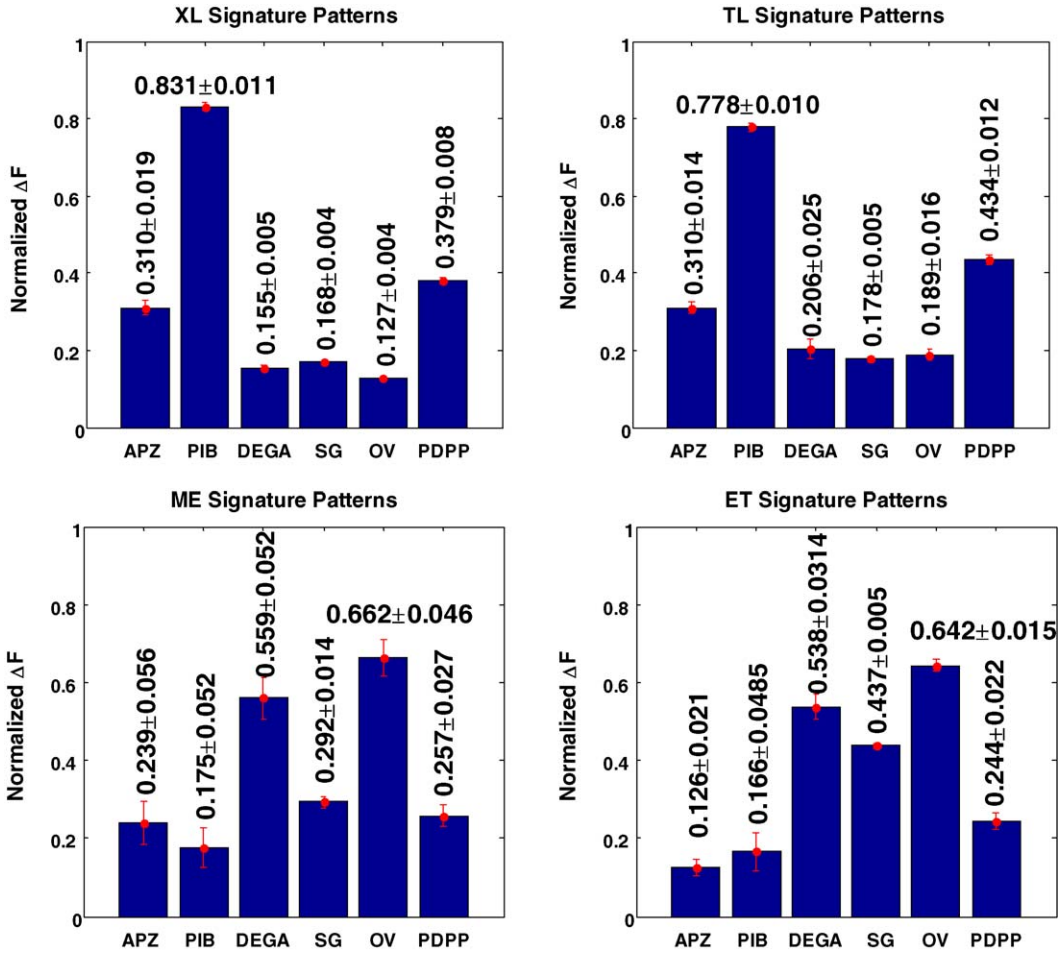

Fig. 5. QCM responses to XL, TL, ME and ET analytes only.

and toluene are very dominant VOCs with very similar response patterns. This is not surprising, as both belong to the same family of benzenes. The responses of the alcohols (ME and ET) do look different than those of benzenes, but are still similar to each other.

To address the problem of overlapping clusters, we discuss two pattern separability algorithms used in tandem with a two-tiered classification procedure, where the dominant VOC is estimated first using an automated classifier, and the secondary VOC is then estimated using a subsequent classifier based on the identification of the dominant VOC.

3.2. Feature extraction and dimensionality reduction for pattern separability

Poor separability problems are common among pattern recognition applications. In general, feature extraction algorithms are employed as a preprocessing step to classification, whose fundamental objective is to obtain the least number of features that carry the most discriminatory information. The general problem of feature extraction can be formulated as one of determining a mapping of the form $\mathbf{y} = f(\mathbf{x})$, or $\mathbf{y} = W^{\mathrm{T}}\mathbf{x}$, that transforms pattern vectors onto a lower dimensional space in which the corresponding feature vectors are better separable. Several well-established techniques have been used (and sometimes misused) for this purpose. For example, the principal component analysis (PCA) is one such popular technique, however, PCA does not take the separability of the patterns into consideration [11]. Therefore, it is strictly a dimensionality reduction technique, rather than a pattern separability based feature extraction technique. The Fisher linear discriminant (FLD) also achieves dimensionality reduction, but by taking a criterion function into consideration: the ratio of intercluster to intracluster distances. The FLD projects the data onto a lower dimensional space where this criterion function is maximized. Consequently, FLD is a feature reduction algorithm that ensures maximum separability of patterns in the transformed space [11]. However, FLD has its limitations. First, regardless of the dimension of the original pattern, the FLD transforms a pattern vector onto a feature vector, whose dimension can be at most $C-1$, where C is the number of classes. This restriction poses a problem in many applications where the original data is of high dimensionality (with many sensors), but few classes. In this case, the originally high dimensional space is forced into a very low $C-1$ dimensional space, and the derived $C-1$ features may just not be sufficient to adequately model the originally high dimensional data distribution. Second, the matrix inversion used in FLD requires that $N-C>d$, where N is the number of training

data and d is the dimension of the pattern vector. This restriction is also a problem, as any application with a large number of classes with respect to available data cannot be analyzed with the FLD approach. However, a modification of FLD originally described in [12], and known as nonparametric discriminant analysis (NDA), removes the above mentioned restrictions by redefining the intercluster distances.

Another technique that can be used for increasing pattern separability is the nonlinear cluster transformation (NCT) proposed in this paper. NCT attempts to increase the intercluster distances while preserving the dimensionality of the pattern vectors. NCT has no limitations in terms of dimensionality, number of classes, or the total number of patterns in the database.

We first describe these two techniques, NDA and NCT, as preprocessing steps for analyzing binary mixture VOC data, followed by a discussion on the two-tiered classification.

4. Methods

4.1. Nonparametric discriminant analysis (NDA)

Consider a multi-class classification problem and let C be the number of classes. For the ith class, let $\{\mathbf{X}_i\}$ be the set of patterns in this class, $\mathbf{m}_i$ be the mean of vectors $\mathbf{x} \in \{\mathbf{X}_i\}$, n_i be the number of patterns in $\{\mathbf{X}_i\}$. Let $\mathbf{m}$ be the mean of all patterns in all C classes. The within scatter matrix $\mathbf{S}_\mathrm{W}$, and between scatter matrix $\mathbf{S}_\mathrm{B}$ are defined as follows:

$$\mathbf{S}_\mathrm{W} = \sum_{i=1}^{C} \sum_{\mathbf{x} \in X_i} (\mathbf{x} - \mathbf{m}_i) \cdot (\mathbf{x} - \mathbf{m}_i)^\mathrm{T},$$

$$\mathbf{S}_\mathrm{B} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{C} \sum_{\mathbf{x} \in X_i} w_{ijx} (\mathbf{x} - \mathbf{m}_{ijx}) \cdot (\mathbf{x} - \mathbf{m}_{ijx})^\mathrm{T} \tag{2}$$

where $\mathbf{m}_{ijx}$ represents the mean of $\mathbf{x}_i$'s k-nearest neighbors from class j, the w_{ijx} represents the weight of the feature vector $\mathbf{x}$ from class i to class j

$$w_{ijx} = \frac{\min(\mathrm{dist}(\mathbf{x}^i_{\mathrm{KNN}}), \mathrm{dist}(\mathbf{x}^j_{\mathrm{KNN}}))}{\mathrm{dist}(\mathbf{x}^i_{\mathrm{KNN}}) + \mathrm{dist}(\mathbf{x}^j_{\mathrm{KNN}})} \tag{3}$$

where $\mathrm{dist}(\mathbf{x}^i_{\mathrm{KNN}})$ is the Euclidean distance from $\mathbf{x}$ to its k-nearest neighbors in class i. In general, if a point belonging to class i is far away in the feature space from the cluster of class j instances, w_{ijx} is a small quantity. If, however, an instance of class i is close to the boundary of class j instances, then w_{ijx} is a large quantity. Hence w_{ijx} is a measure of the proximity between pattern $\mathbf{x}$ of class i and class j patterns as a whole. We note that $\mathbf{S}_\mathrm{W}$ is a measure of the intracluster distances, and $\mathbf{S}_\mathrm{B}$ is a measure of the intercluster distances. We should add that the main difference between FLD and the NDA is the definition of the $\mathbf{S}_\mathrm{B}$. While the definition of $\mathbf{S}_\mathrm{W}$ remains the same in both approaches, in FLD $\mathbf{S}_\mathrm{B}$ is defined as

$$\mathbf{S}_\mathrm{B} = \sum_{i=1}^{C} N_i (\mathbf{m} - \mathbf{m}_i) \cdot (\mathbf{m} - \mathbf{m}_i)^\mathrm{T} \tag{4}$$

where N_i is the number of data instances in class i. It is useful to note that $\mathbf{S}_\mathrm{B}$ is the sum of C matrices each of which has a rank of at most 1. These matrices calculate a measure of difference between the overall mean $\mathbf{m}$ of all data and individual cluster means $\mathbf{m}_i$, and hence $\mathbf{S}_\mathrm{B}$ represents the intercluster distances. Since only $C-1$ of these matrices are independent, $\mathbf{S}_\mathrm{B}$ is of rank $C-1$ or less, which is the source of the restriction mentioned above. The revised definition of $\mathbf{S}_\mathrm{B}$ in (4) removes this restriction by forcing $\mathbf{S}_\mathrm{B}$ to be a full-rank matrix (of rank d for d-dimensional data). The transformation, the projection from the original feature space onto a lower dimensional feature space, can then be expressed as

$$\mathbf{y} = \mathbf{W}^\mathrm{T} \cdot \mathbf{x} \tag{5}$$

where the column vector $\mathbf{y}$ is the feature vector in the projected space corresponding to pattern $\mathbf{x}$. The optimum matrix $\mathbf{W}$ is obtained by maximizing criterion function:

$$J(\mathbf{W}) = \frac{|\mathbf{W}^\mathrm{T} \mathbf{S}_\mathrm{B} \mathbf{W}|}{|\mathbf{W}^\mathrm{T} \mathbf{S}_\mathrm{W} \mathbf{W}|} \tag{6}$$

The columns of $\mathbf{W}$, denoted as $\mathbf{w}_i$, that maximize $J(\mathbf{W})$ are then the eigenvectors that correspond to the largest eigenvalues in the generalized eigenvalue equation [11]:

$$\mathbf{S}_\mathrm{B} \mathbf{w}_i = \lambda_i \mathbf{S}_\mathrm{W} \mathbf{w}_i \tag{7}$$

For nonsingular $\mathbf{S}_\mathrm{W}$, Eq. (7) can be written as

$$\mathbf{S}_\mathrm{W}^{-1} \mathbf{S}_\mathrm{B} \mathbf{w}_i = \lambda_i \mathbf{w}_i \tag{8}$$

From Eq. (8), we can directly compute the eigenvalues λ_i and the eigenvectors $\mathbf{w}_i$, constituting the columns of the $\mathbf{W}$ matrix, which can then be used to obtain the transformed instances $\mathbf{y}$ in the new feature space.

4.2. Nonlinear cluster transformation (NCT)

NCT is a three step procedure: in the first step, reduction of intracluster distances is achieved by eliminating the outliers. In the second step, the desired cluster separation is obtained by a simple translation of each cluster along an optimal direction. This step, in essence, generates training data pairs for determining the NCT function for the third step. In the last step, the data generated in step two is used to train a generalized regression neural network (GRNN) to approximate the function mapping between original clusters and the translated clusters. The feature vectors are then input to a classifier of choice. The details of these steps are explained below.

4.2.1. Outlier removal

The patterns in each class i in the training database are first normalized according to

$$\mathbf{x} = \frac{\mathbf{x}}{\sqrt{\sum_{k=1}^{d} (x^k)^2}} \tag{9}$$

where x^k is the kth element of the pattern $\mathbf{x}$, and d is the dimensionality of the patterns. Outlier removal is performed next,

based on the Mahalanobis distances of patterns from the cluster centers. For each cluster i, the Mahalanobis distance of pattern $\mathbf{x}$ in class i is computed as

$$M_{\mathrm{D}} = (\mathbf{x} - \mathbf{m}_i)^{\mathrm{T}} \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i) \quad \mathbf{x} \in \{\mathbf{X}_i\} \tag{10}$$

where $\mathbf{C}_i$ is the covariance matrix of instances of the ith class, and $\mathbf{m}_i$ is the mean of this population. M_{D} can be used as a measure of dispersion within the cluster. A suitable threshold is chosen based on the data, and instances with an M_{D} larger than this threshold are removed.

4.2.2. Cluster translation

This step addresses the problem of closely packed and possibly overlapping clusters. The idea is to translate the clusters appropriately in order to physically separate them. Conceptually, all clusters are thought of as like charged particles: the magnitude and direction of the translation vector are then derived using the concept of a repulsive force exerted by each cluster i on other clusters.

Consider a two-class problem with (possibly) overlapping clusters, whose centers are located at $\mathbf{m}_1$ and $\mathbf{m}_2$. The distance between these two clusters can be increased if class I patterns are translated along the vector $\mathbf{S}_1 = -(\mathbf{m}_2 - \mathbf{m}_1)$, and class II patterns are translated along $\mathbf{S}_2 = -\mathbf{S}_1 = (\mathbf{m}_2 - \mathbf{m}_1)$. This idea can be extended to multi-class problems of arbitrary dimensionality, where patterns of class C_i can be translated along $\mathbf{S}_i$, where the optimal direction $\mathbf{S}_i$ can be computed as

$$\mathbf{S}_i = -\sum_{j \neq i}^{C} (\mathbf{m}_j - \mathbf{m}_i) \tag{11}$$

and where $\mathbf{m}_i$ and $\mathbf{m}_j$ are the cluster centers of class i and class j, respectively, and C is the number of classes. The resultant translation vector for class i is $\mathbf{S}_i = -\mathbf{M}_i$, where

$$\mathbf{M_i} = \sum_{j \neq i} (\mathbf{m}_j - \mathbf{m}_i) \tag{12}$$

All patterns in class i are moved along the direction of $-\mathbf{M}_i$, and the translated patterns can be obtained by

$$\mathbf{x}_{\mathbf{S}_i} = \mathbf{x}_i + \left(-\frac{\mathbf{M}_i}{\|\mathbf{M}_i\|} \right) \mathrm{dist}_i \tag{13}$$

where $\mathbf{x}_i$ is a pattern from class i, $\mathrm{dist}_i = 1/|\mathbf{m} - \mathbf{m}_i|$ is a normalizing constant that controls the amount of translation, $\mathbf{X}_{\mathbf{S}_i}$ is the new location of the pattern $\mathbf{x}_i$, $\mathbf{m}_i$ is the mean of all class i instances, and $\mathbf{m}$ is the mean of all instances from all classes. If the overlap in the data from different classes is high, dist_i should be correspondingly high. For data with up to three dimensions, this parameter can be determined simply by plotting the data and determining how much translation is necessary. For larger dimensional data, it can be determined by trial and error or prior experience. It is straightforward to show mathematically that these translation directions maximize intercluster distances [13]. Note that $\mathbf{S}_i$ points in the opposite direction of the resultant vector that combines the cluster center of class i to the centers of all other classes, that is, it points away from all other clusters. The procedure is conceptually illustrated in Fig. 6.

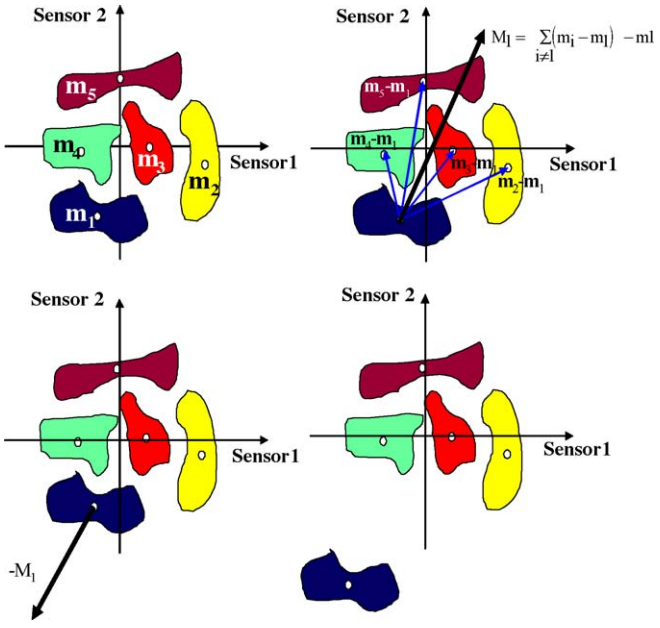

Fig. 6. Nonlinear cluster transformation.

The cluster transformation described here can also be expressed in a matrix form. Let $i = 1, 2, \ldots, C$, where C is the number of classes, $n = 1, 2, \ldots, N_i$ where N_i is the number of patterns in class i, $\mathbf{x}_n^i$ be the d-dimensional nth pattern of the ith class, and $\mathbf{y}_n^i$ be the corresponding d-dimensional pattern after translation. Then:

$$\begin{pmatrix} \mathbf{y}_1^i \\ \mathbf{y}_2^i \\ \vdots \\ \vdots \\ \mathbf{y}_{N_i}^i \end{pmatrix}_{N_i \times d} = \mathrm{dist}_i \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \\ & \vdots & \\ & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{N_i \times C} \begin{pmatrix} \mathbf{m}_1 - \mathbf{m}_i \\ \mathbf{m}_2 - \mathbf{m}_i \\ \vdots \\ \vdots \\ \mathbf{m}_C - \mathbf{m}_i \end{pmatrix}_{C \times d} + \begin{pmatrix} \mathbf{x}_1^i \\ \mathbf{x}_2^i \\ \vdots \\ \vdots \\ \mathbf{x}_{N_i}^i \end{pmatrix}_{N_i \times d} \tag{14}$$

This equation can be implemented on the training data sets to generate a second dataset that can be used to train a neural network to learn the overall transformation function.

4.2.3. Function mapping

In order to translate each cluster away from each other, the correct class information is required, which obviously is not available for a test pattern. We therefore need to *learn* how to translate patterns without knowing the class information. This problem can be thought of as a function approximation prob-

lem, where the function to be approximated is a function that maps d-dimensional original patterns to their new locations. A generalized regression neural network (GRNN) was used to accomplish this function approximation. GRNNs can be thought of as a special case of radial basis function neural networks (RBFNN). GRNNs do not require iterative training, and they can approximate any arbitrary multidimensional function defined between a set of input and output vectors. GRNN is based on the theory of nonlinear regression analysis, commonly used as a statistical function estimation scheme. For brevity and due to their widespread use, GRNN architecture is not reviewed here, and interested readers are referred to [14].

4.3. Two-tier classification

While both of the above described techniques address the pattern separability problem, the presence of a dominant VOC still poses a considerable challenge, as the response of the secondary VOC may be completely masked. In such cases, the pattern separability approaches work best, when they are assisted with a two-tiered classification, where the dominant VOC is identified first, and the secondary VOC is determined based on the information obtained in the first step. A neural network based automated classifier is first trained to identify one of the five dominant VOCs present in the mixture. Once the dominant VOC is identified, one of five additional classifiers is used, each of which is specifically trained to recognize the secondary VOCs in a mixture of the given dominant VOC. Fig. 7 illustrates the overall schematic of this divide and conquer type approach.

The inputs to a particular secondary network are the same preprocessed inputs for the dominant VOC networks. However, only those instances identified to belong to the dominant class c are used as inputs to the secondary VOC network for class c. For instance, once a particular data instance is identified to include toluene as its dominant VOC, the same data instance (preprocessed either by NDA or NCT) is the input to the secondary VOC network for toluene.

The networks shown in Fig. 7 were all multilayer perceptron type neural networks, whose widely reported success on a variety of pattern recognition applications has made them a popular choice.

5. Results

The gas-sensing data and known corresponding VOC mixtures – with no preprocessing for pattern separability – were first used to train various architectures of multilayer perceptron (MLP), as well as radial basis function (RBF) neural networks. These networks failed to converge – let alone perform satisfactorily – for all reasonable architectures and learning parameters, even for identification of the dominant VOC only. The networks were unable to learn the decision boundaries separating the clusters formed by different VOCs. This behavior was traced to the problem of closely packed or possibly overlapping clusters in the six-dimensional space, as mentioned above and illustrated in Figs. 3–5.

Both NDA and NCT were first performed on the problem of dominant VOC identification, a five-class (OC, ET, XL, TL and TCE), six-feature classification problem. The 384-pattern database (16 measurements for each of the 24 mixtures) was partitioned into two equal parts of 192 instances, **PS** for training and evaluating pattern separability, and **CL** for training and evaluation of the two-tiered final classification.

For NCT, the database **PS** was also partitioned for training and evaluation of pattern separability. The training partition and its translated target vectors were used to train the GRNN to learn the functional mapping required to obtain the desired transformation. The GRNN had 6 input nodes, 192 hidden layer nodes (receptive fields), and 5 output nodes. Various σ values (spread constant) were used and the best separation was obtained for σ in the 0.06–0.1 range. The trained GRNN was then evaluated on the validation partition of the dataset **PS**. We note that this is not the validation of the final classification performance, but rather a figure of merit for the ability of the algorithm to separate the patterns.

For training and testing the classification capability of a neural network using the NDA or NCT processed data, the dataset **CL**, consisting of the remaining 192 six-dimensional measure-

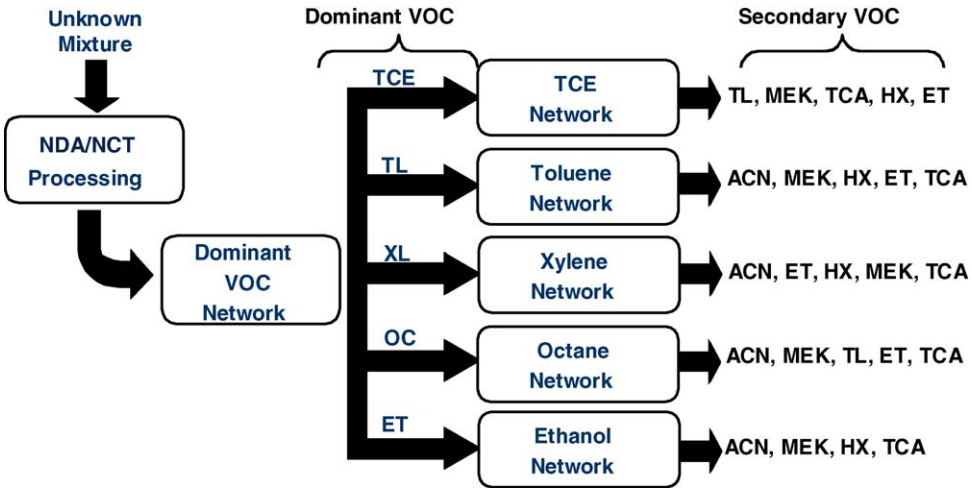

Fig. 7. Two step approach for identification of individual VOCs within a mixture.

ments, was further partitioned into two: $\mathbf{TR}_{CL}$ consisting of 48 patterns (2 measurements per mixture—for training) and $\boldsymbol{E}_{CL}$ consisting of 144 patterns (6 measurements per mixture—for final evaluation of classification performance). A two hidden layer MLP with $6 \times 40 \times 20 \times 5$ architecture was trained using the $\mathbf{TR}_{CL}$ data, which easily converged to a mean square error minimum of 0.01. Repeating with σ values between 0.06 and 0.1, a correct classification performance of 98% was achieved on the $\boldsymbol{E}_{CL}$ for identification of dominant VOCs using NCT, and 86.4% using NDA. This is in comparison to complete failure of convergence with unprocessed data.

Once the dominant VOC was identified, a separate MLP was trained for each of the five secondary VOCs, creating the two-tier classification system. A similar procedure was repeated for the identification of secondary VOCs, training $6 \times 20 \times 5$ MLP networks with preprocessed patterns for each dominant VOC. The overall performance of the secondary VOC networks was 97% with NCT and 82% with NDA.

6. Conclusions and discussion

Two pattern separability techniques, NDA and NCT, have been applied to mixture VOC data. NDA aims to maximize the intercluster to intracluster distance ratios, whereas NCT tries to increase the intercluster distances while keeping intracluster distances constant. Preprocessing allowed improved performances of subsequent classification algorithms, and in fact, made training possible for the VOC database.

For identification of binary mixtures of VOCs, where the response to a secondary VOC may be masked by that of a dominant VOC, we propose a two-tiered classification procedure. The dominant VOC is identified first, and based on this information a second level of classifiers are used – one for each dominant VOC – to identify the secondary VOCs. Attempting to identify both components at once had earlier proven to be impractical.

Both NDA and NCT had satisfactory performances for the identification of dominant VOCs, but NCT performed better for the identification of secondary VOCs. The advantage of NDA is that it can provide dimensionality reduction, if such an operation is necessary for the particular application. However, NDA is also computationally more expensive than NCT, which is designed to preserve the dimensionality of the problem.

Acknowledgments

This work is supported by the New Jersey Water Resources Research Institute administered by Cook College at Rutgers University, NJ; and the National Science Foundation under grant No: ECS-0239090. Experimental sections of this work were conducted at Microanalytical Instrumentation Center (MIC) at Iowa State University, Ames IA. Authors greatly acknowledge the assistance of Dr. Ruth Shinar and Dr. Marc Porter at MIC.

References

[1] T.C. Pearce, et al., Handbook of Machine Olfaction, Electronic Nose Technology, Wiley, 2003.
[2] A.K. Srivastava, Detection of volatile organic compounds (VOCs) using SnO_2 gas-sensor array and artificial neural network, Sens. Actuators B: Chem. 96 (2003) 24–37.
[3] M. Sriyudthsak, A. Teeramongkolrasasmee, T. Moriizumi, Radial basis neural networks for identification of volatile organic compounds, Sens. Actuators B: Chem. 65 (2000) 358–360.
[4] R. Polikar, R. Shinar, V. Honavar, L. Upda, M.D. Porter, Detection and identification of odorants using an electronic nose, in: Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT, 2001, pp. 3137–3140.
[5] A. Ortega, S. Marco, T. Sundic, J. Samitier, New pattern recognition systems designed for electronic noses, Sens. Actuators B: Chem. 69 (2000) 302–307.
[6] M. Penza, G. Cassano, Application of principal component analysis and artificial neural networks to recognize the individual VOCs of methanol/2-propanol in a binary mixture by SAW multi-sensor array, Sens. Actuators B: Chem. 89 (2003) 269–284.
[7] C. Di Natale, F.A.M. Davide, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Gopel, Composed neural network for the recognition of gas mixtures, Sens. Actuators B: Chem. 25 (1995) 808–812.
[8] E. Llobet, R. Ionescu, S. Al-Khalifa, J. Brezmes, X. Vilanova, X. Correig, N. Barsan, J.W. Gardner, Multicomponent gas mixture analysis using a single tin oxide sensor and dynamic pattern recognition, IEEE Sens. J. 1 (2001) 207–213.
[9] W.H. King, Piezoelectric sorption detector, Anal. Chem. 36 (1964) 735–1739.
[10] H. Wohltjen, Mechanism of operation and design considerations for surface acoustic wave device vapor sensors, Sens. Actuators 5 (1984) 307–325.
[11] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley Interscience, New York, NY, 2001 (Chapter 3).
[12] K. Fukunaga, Statistical Pattern Recognition, 2nd ed., Academic Press, New York, NY, 1990.
[13] R. Polikar, L. Udpa, S.S. Udpa, Nonlinear cluster transformations for increasing pattern separability, in: Proceedings of the International Joint Conference on Neural Networks, Washington, DC, USA, 1999, pp. 4006–4011.
[14] D.F. Specht, A general regression neural network, IEEE Trans. Neural Netw. 2 (1991) 568–576.

Biographies

Robi Polikar received his BSc degree in electrical engineering from Istanbul Technical University in 1993, and his MS and PhD degrees, both in biomedical engineering and electrical engineering, from Iowa State University, in 1995 and 2000, respectively. Since 2001, he has been with the Department of Electrical and Computer Engineering at Rowan University, New Jersey. His current research interests include computational intelligence, signal processing, pattern recognition and data fusion particularly in applications to biological and chemical signals.

Kauser Jahan received her BSc degree in civil engineering from Engineering University in Dhaka in 1983, her MS degree in environmental engineering from University of Arkansas in 1987, and her PhD degree from University of Minnesota in 1993. She has been with the Department of Civil and Environmental Engineering at Rowan University, New Jersey since 1996. Her primary experience is in fate and transport of organic pollutants in soil and water.