Microchim Acta 156, 183–207 (2007)
DOI 10.1007/s00604-006-0623-9

Microchimica Acta
Printed in Austria

Review

Data analysis for electronic nose systems

Simon M. Scott, David James, and **Zulfiqur Ali***

School of Science and Technology, Applied Science Department, University of Teesside, Middlesbrough, TS1 3BA, United Kingdom

Received July 6, 2005; accepted April 6, 2006; published online September 29, 2006

Abstract. Electronic noses (e-noses) employ an array of chemical gas sensors and have been widely used for the analysis of volatile organic compounds. Pattern recognition provides a higher degree of selectivity and reversibility to the systems leading to an extensive range of applications. These range from the food and medical industry to environmental monitoring and process control. Many types of data analysis techniques have been used on the data produced. This review covers aspects of analysis from data normalisation methods to pattern recognition and classification techniques. An overview of data visualisation such as non-linear mapping and multivariate statistical techniques is given. Focus is then on the use of artificial intelligence techniques such as neural networks and fuzzy logic for classification and genetic algorithms for feature (sensor) selection. Application areas are covered with examples of the types of systems and analysis methods currently in use. Future trends in the analysis of sensor array data are discussed.

Key words: Electronic noses; sensor arrays; pattern recognition; sensor selection.

Instrumental methods of determining volatiles, such as gas chromatography – mass spectrometry (GC/MS), are expensive and require trained personnel. As a result there has been a drive to establish a device for rapid, inexpensive analysis of volatile organic compounds (VOC) that do not require specialist technicians. Chemical sensors are transducers that incorporate a chemical detection layer and transform a chemical interaction into a measurable signal; they offer rapid and inexpensive detection of VOC with simple interpretation. The ideal gas sensor would exhibit reliability, robustness, sensitivity, selectivity and reversibility. High selectivity with high reversibility is difficult to attain, either a compromise is necessary or the detection layer of the sensor must be regenerated.

An e-nose is an instrument consisting of an array of reversible but only semi-selective gas sensors coupled to a pattern recognition algorithm. The selectivity of the instrument is achieved through the application of pattern recognition techniques to the responses from the senor array. Persaud and Dodd [1] first reported the design of an e-nose to analyse a pattern of responses from a sensor array to differentiate between members of a pre-defined set of volatiles.

Sensor types used in e-noses have been many and varied. Conducting polymer sensors (CP), piezoelectric – surface acoustic wave (SAW), thickness shear mode (TSM), metal oxide semiconductor (MOS), metal oxide semiconductor field effect transistor (MOSFET), electrochemical (EC), Pellistor and optical sensor arrays are reviewed by Albert et al. [2] and James et al. [3]. Assessments of chemical sensors may be found in [4–8]. General books and reviews [9–11] cover aspects of sensor technology. Headspace mass spectrometry (HS-MS) systems (MS based e-noses) consist of introducing volatile components present in the HS

* Author for correspondence. E-mail: z.ali@tees.ac.uk

of a sample without prior chromatographic separation into the ionization chamber of a mass spectrometer. The spectrum resulting from simultaneous ionization and fragmentation of the mixture of molecules introduced constitutes a pattern that is characteristic of the product under analysis [12, 13]. Most applications for e-noses concentrate on four major areas; food [14–19], medical diagnosis [20, 21], environmental monitoring [22–26] and bioprocess control [27–30].

Data acquisition and feature extraction

Data acquisition is the first step for data analysis; sensors collect the data and convert it into an electrical signal pattern that is more suitable for computer analysis. This step often causes the difficulty in the classification problem as the characteristics and limitations of the transducer may limit or distort the available information. The output is a pattern vector, in pattern space. The pattern vector is passed into the second stage, the feature extractor. Feature extraction is the use of one or more transformations of the input features to produce new salient features. Feature extraction may be regarded as a dimensionality reduction process; data is converted from pattern space into feature space. Features should be easily evaluated, there are two kinds, the first have a clear physical meaning, the second have not and are called mapping features. In the context of e-noses the features that are usually considered are the maximum sensor responses, but some groups have used temporal data or mapping features such as constructs of gradient information. Data pre-processing can affect the classifier, there are no general guidelines to determine the appropriate data pre-processing technique. Concentration may have a scaling effect on the patterns of the sensors. While a pattern recognition algorithm normally examines the differences between the patterns, a scaling effect may mask the interrelations between them. Normalization is normally used to remove this effect. Table 1 shows several standard data normalisation techniques [10, 11, 31].

Table 1. Standard data normalisation techniques

Normalisation	Equation	
Relative scale$_1$	$X_{ij} = \dfrac{X_{ij}}{\max(\boldsymbol{X})}$	(1)
Relative scale$_2$	$X_{ij} = \dfrac{X_{ij}}{\max(\mathbf{x}_j)}$	(2)
Relative scale$_3$	$X_{ij} = \dfrac{X_{ij}}{\max(\mathbf{x}_r)}$, $\mathbf{x}_r$ is a $1 \times p$ reference response	(3)
Relative scale$_4$	$X_{ij} = \dfrac{X_{ij}}{\lVert \mathbf{x}_i \rVert}$	(4)
Background subtraction	$\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_b$, where $\mathbf{x}_b$ is the $1 \times p$ blank response	(5)
Signal average	$X_{ij} = \dfrac{1}{N}\sum_{k=1}^{N} X_{ij}^k$, N replicates of data	(6)
Auto scale	$X_{ij} = \dfrac{X_{ij} - \mathbf{x}_j}{\sigma_{\mathbf{x}_j}}$	(7)
Range scale$_1$	$X_{ij} = \dfrac{X_{ij} - \min(\mathbf{x}_j)}{\max(\mathbf{x}_j) - \min(\mathbf{x}_j)}$	(8)
Range scale$_2$	$X_{ij} = \dfrac{2(X_{ij} - \min(\mathbf{x}_j))}{\max(\mathbf{x}_j) - \min(\mathbf{x}_j)} - 1$	(9)
Baseline subtraction	$X_{ij} = X_{ij} - X_{1j}$	(10)

$\boldsymbol{X}$ is the feature matrix of n samples from p sensors, X_{ij} is the i^{th} sample of the j^{th} sensor, $\mathbf{x}_j$ contains all n response samples for sensor j, $\mathbf{x}_i$ contains all p responses for sensors at the i^{th} sample. Relative scale$_1$ gives a global compression of values with a maximum value of 1, Relative scale$_2$ compresses values per feature with a maximum value of 1, Relative scale$_3$ is relative to a reference, and Relative scale$_4$ is the Euclidean distance norm. Relative scaling is often used in qualitative applications. Signal average uses the mean of several runs to smooth the noise present in the signal or the mean of several of the same sensor types in an array. Mean-centre sets the mean of the data to the origin. Auto scale sets the mean at the origin and the variance within the data to 1, often used when responses are on different magnitude scales. Range scale$_1$ and range scale$_2$ set the limits at [0, 1] and [−1, 1] respectively. Background subtraction attempts noise correction by removing the response from a blank sensor. A baseline subtraction removes the base reading of a sensor, and is often used in temporal data collection. Applications that use temporal data produce a large number of measurements; typically these are reduced by pre-processing before pattern recognition is attempted.

Curse of dimensionality

The notion of the curse of dimensionality was introduced by Bellman [32] as a result of studies in adaptive control processes. The problem stems for the number of data points needed to adequately represent a data set with a high number of features; it is quite possible that within high dimensional data, clusters exist in separate sub-spaces. All classifiers can suffer

from the curse [33]. The only practical way to beat the curse is to apply prior knowledge on the data [34], or to carefully select the minimum number of features to adequately represent the problem.

Feature selection

Feature selection is the process of identifying the most effective subset of the original features to use in the classification process that lead to the smallest classification error. Blum and Langley [35] classify feature selection techniques into three basic approaches. In the first, known as the embedded approach, features are added or removed in response to prediction errors of a simple embedded classifier. The second are filter methods and work independently to remove features without knowing the effect on the classification algorithm, typical linear methods used are principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA) [36, 37]. The third are wrapper methods and evaluate candidate feature sets using a classification algorithm on the training data. The feature subset selection algorithm conducts a search for a good subset using the classifier as part of the evaluation function [38].

Similarity

The idea of similarity is fundamental to the definition of a cluster of objects. Similarity is usually measured by a distance function defined on pairs of patterns. As there are a wide variety of feature types and scales, the distance measure should be chosen carefully. It is most common to calculate the dissimilarity between two patterns using a distance measure defined on the feature space. The most popular metric for continuous features is the Euclidean distance.

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{d} (\mathrm{x}_{ik} - \mathrm{x}_{jk})^2 \right)^{\frac{1}{2}} \quad (1)$$

$$= \|\mathbf{x}_i - \mathbf{x}_j\| \quad (2)$$

The Euclidean distance has an intuitive appeal and is commonly used to evaluate the proximity of objects in two and three-dimensional space. The drawback of this metric is the tendency of the largest scaled variables to dominate the others. Solutions to this problem include normalisation of the continuous features or other weighting schemes. Linear correlation among features can also distort distance measures. This distortion may be alleviated by using the squared Mahalanobis distance

$$d_M(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) \sum^{-1} (\mathbf{x}_i - \mathbf{x}_j)^T \quad (3)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are assumed to be row vectors, $(\mathbf{x}_i - \mathbf{x}_j)^T$ is the transpose of the $(\mathbf{x}_i - \mathbf{x}_j)$ vector and $\sum^{-1}$ is the inverse of the sample covariance matrix of the patterns or the known covariance matrix of the pattern generation process; d_M assigns different weights to different features based on their variances and pair wise linear correlations.

Pattern recognition techniques

The data produced by an e-nose result in a set of semi-independent variables (the sensor array output) and a set of dependent variables (odour classes). There are two distinct types of pattern recognition, those used as exploratory and confirmatory techniques. Two other types of analysis used are for feature selection and sensor drift counteraction techniques.

Most exploratory techniques are unsupervised, whilst classifiers are supervised.

- During supervised learning the data vectors are tagged with a descriptor, the classes are learned and grouped according to their description. After learning is complete an unknown vector may be classified using the relationships found *a priori* from the known vectors.
- In unsupervised learning there are no descriptors, the classes are learnt based on some form of similarity measure alone.

Statistical methods are parametric as they assume the data may be described in terms of probability density functions and include PCA, hierarchical cluster analysis (HCA), principal component regression (PCR), discriminant factor analysis (DFA), analysis of variance between groups (ANOVA), partial least squares (PLS) regression method, principal components regression (PCR), soft independent modelling class analogy (SIMCA) [39] and clustering algorithms such as k-means.

Artificial intelligence (AI) techniques are generally non-conventional, intuitive approaches for problem solving often biologically inspired and may be split into three sub-groups.

- Artificial neural networks (ANN), including multi layer perceptrons (MLP) and radial basis function

networks (RBF), self organising maps (SOM), learning vector quantization (LVQ) and self organising competitive systems such as the adaptive resonance theory (ART) or growing cell family of algorithms.
– Fuzzy logic and fuzzy rules based algorithms.
– Genetic algorithms (GA) are used for feature selection.

Graphical methods for exploratory analysis

Sensor arrays may generate a large volume of high-dimensional data, it is often a challenge to extract useful information from the data to solve the problem under investigation. Graphical methods are a simple exploratory way of analysing data, some methods may plot the high-dimensions produced, whilst others reduce the data to two or three dimensions for visual analysis.

Bar charts

Bar charts are a simple and useful way of visualising the sensor response patterns that are produced by a sensor array. The sensor responses are plotted as a height value with a small width, next to each other. Individual or the mean of several samples are plotted for each analyte. An estimation of the effectiveness and composition of the array may then be made based on the patterns produced. Walmsley et al. [40] used bar charts of the response of 4 MOS sensor array to show that the patterns produced were different for each of 6 analytes. McAlernon et al. [41] used 3D bar charts, PCA and multivariate analysis of variance (MANOVA) to interpret the frequency change patterns of an array of 8 TSM sensors for *o*-xylene, toluene, dodecane and tetradecane. Park and Zellers [42] used bar charts to show the relative response patterns of polymer coated SAW sensors to toluene and trichloroethylene. Groves and Zellers [43, 44] analysed solvent vapours in breath and ambient air with an array of 4 polymer coated SAW sensors. Time response curves and bar charts were used to show the relative response patterns for 16 solvents. The bar charts provided a visual indication of the discriminating capability of the array to the various solvents.

Polar plots

Polar plots or radar plots a.k.a Star plots [45] display multivariate dimensional data in two dimensions for exploratory analysis. Axes for each feature radiate from the origin at equal angles with the magnitudes of the features joined by straight lines. Figure 1 shows a typical plot for a 6 sensor array. Some pre-processing of the data may exaggerate small difference in shapes so that they may be more easily seen. For example a reference data set may be subtracted from the feature vectors and rescaled to emphasise the differences.

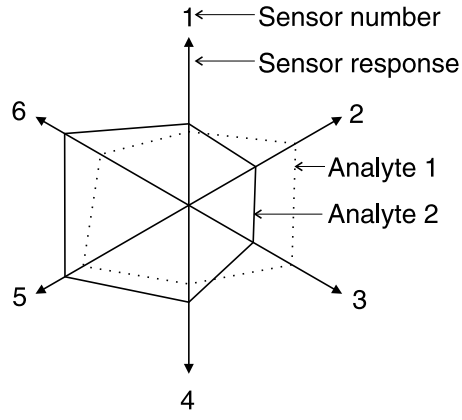

Fig. 1. A polar plot of sensor response data

Brezmes et al. [46] used polar plots to compare both the difference between repeated measurements and measurements of different aromatic species; cinnamon, red pepper, thyme, pepper and nutmeg. The method showed large differences in the shapes for different species, and possible poor reproducibility by smaller differences between repeated runs of the same groups. Jonsdottir and co-workers [47] characterised the flavours of ripened cod roe by polar plots utilising the results of a sensory panel. From this an array of 4 EC sensors were used to examine the same cod roe samples. Gan et al. [48, 49] characterised responses to palm olein and vegetable oils from a virtual array of SAW sensors, based on a single sensor responding to the analyte being desorbed from a GC column trap, plotting the raw frequency shift as amplitude and the time as the angle. PCA scores were plotted to demonstrate the separation achieved but no classification algorithm was tried.

Hierarchical cluster analysis (HCA)

HCA techniques attempt to separate data into specific groups [50], based on a similarity measure. Initially each data point represents its own cluster, then the threshold for the decision when to declare two or more objects to be members of the same cluster is lowered

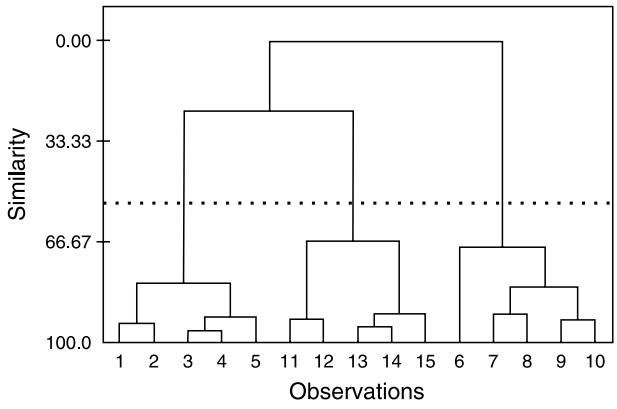

Fig. 2. Dendogram illustrating HCA data clustering

incrementally. As a result more and more objects are linked together and aggregated into larger and larger clusters of increasingly dissimilar elements. Finally, all objects are joined together. The result of hierarchical clustering methods may be displayed as a dendrogram, as shown in Fig. 2; the vertical axis denotes the similarity. With classification set at a similarity level of 50%, the data has been classified into 3 groups. There are many types of similarity linkage, the three most common are single linkage (nearest neighbour), the distance between two clusters is determined by the distance of the two closest objects in the clusters. Complete linkage (furthest neighbour), the distances between clusters is determined by the greatest distance between any two objects in the different clusters. Group average, the distances between clusters is determined by the mean distance of objects in different clusters.

Sammon mapping

Sammon mapping is a nonlinear mapping (NLM) technique [51] that attempts to preserve the structure of the data in the projected space from the original high dimensional space by maintaining the distances between points under projection [52, 53]. The distance metric usually employed is the Euclidean, although others have been used [54, 55]. If the distance between two points i and j in the input space of vector $\mathbf{x}$ is $d_{ij}^* = d(x_i, x_j)$ and the distance between the points in projected space of vector $\mathbf{y}$ is $d_{ij} = d(y_i, y_j)$. Sammon suggested looking for values of $\mathbf{y}$ to minimise an error function called the mapping stress E:

$$E = \frac{1}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} \quad (4)$$

Sammon used a method of steepest descent for (approximate) minimisation of E but there are many local minima on the error surface and it is unavoidable for the algorithm to become stuck. The algorithm is usually run several times with different initial configurations and the outcome with the lowest stress chosen. There is a high computational load, which is $O(n^2)$. At every iteration $n(n-1)/2$ distances and error derivatives must be calculated. As n the number of patterns increases the computational requirements increase quadratically.

Sammon mapping applications

Persaud et al. [56] used Sammon mapping to help determine the clusters of detected volatiles found in pig slurry from animals fed on two different diets, and whether biscuits had become slightly rancid on ageing from the normalised responses of 20 CP sensors. Byun et al. [57] applied Sammon mapping to data from an array of 20 CP sensors to show the clustering obtained from 5 test alcohols using Euclidean distances. Persauds' group [58] used bar charts of sensor responses, PCA scores and Sammon mapping to describe results of environmental monitoring of the MIR space station from an array of 20 CP sensors operating at 40 °C. The system was able to provide information of transient changes of the atmosphere at 30 second intervals, including a retrospective analysis to detect an ethylene glycol leak of the cooling system.

Other graphical methods

Bourgeois et al. [59] used a graph of sensor responses against the sample number in a continuous wastewater sampling system to show the dilution effect of rain on the samples. Biswas et al. [60] used an array of 5 SAW sensors to differentiate between fresh and oxidised vegetable oils. Plots of peak frequency change against oxidisation time (days) showed that the system could detect the difference between the oils and also could be utilised as an analytical tool for following the breakdown of vegetable oils. Nonaka et al. [61] used a preconcentration system with an array of 6 MOS sensors, analysing data by multiple linear regression, receiver-operating characteristic (ROC) plots and organoleptic tests showed that oral malodour intensity may be evaluated by an e-nose.

Principal component analysis

Principal component analysis (PCA) a.k.a. the Karhunen-Lóeve expansion or singular value decomposition (SVD) is an unsupervised multivariate procedure and is a well known linear data compression and feature extraction technique. The scores produced may be plotted in two or three dimensions to inspect the data. PCA derives new, uncorrelated variables that are linear combinations of the original variable set ordered by reducing variability. It is mainly used to reduce the dimensionality of a data set while retaining as much information as possible by eliminating the lowest ranking variables. PCA is a simple and fast method but remains a linear approach therefore any non-linear correlation between variables will not be retained.

If the original feature vectors are held in a matrix $\boldsymbol{X}$, they may be approximated by a product of two small matrices – the score and loading matrices, such that, $\boldsymbol{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$ where $\boldsymbol{X}$ is the original data matrix of n rows and k columns (features), $\mathbf{T}$ is the scores matrix (a rows and n columns – the number of principal components) and $\mathbf{P}$ is the weight or loading matrix with k columns and a rows, $\mathbf{P}^{\mathrm{T}}$ is the transpose of the $\mathbf{P}$ matrix (k rows, a columns). For each score variable $\mathbf{t}$, the influence of the original variables is the corresponding loading profile, $\mathbf{p}$. This is a direct link between the scores $\mathbf{T}$, and the original $\boldsymbol{X}$-variables and is very useful when interpreting scatter plots of scores and loadings.

The space is analysed and a low-dimensional hyperplane that best summarize all the variation in $\boldsymbol{X}$, in terms of least squares is found as shown in Fig. 3. The loading values (weight) for each PC component are the cosine of the angle between the principal component direction and each of the original coordinate axes. The smaller the angle between the vectors the higher the correlations between the features, highly correlated features provide little additional information. Uncorrelated features are orthogonal to each other. Loadings in the origin of the coordinate system represent unimportant features for that component. The length of the feature with respect to the axes gives the proportion of the eigenvectors that the feature represents. The part of $\mathbf{X}$ that is not explained by the model forms the residuals ($\mathbf{E}_{\mathrm{nk}}$) and is the distance between each point in K-space and its point on the plane. The scores, loadings and residuals together describe all of the variation in $\mathbf{X}$.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}} + \ldots + \mathbf{E} \qquad (5)$$

The loadings ($\mathbf{P}$) are ranked in the order of the largest eigenvectors of ($\mathbf{X}^{\mathrm{T}}\mathbf{X}$) and the score vectors ($\mathbf{T}$) are ranked in the order of the largest eigenvectors of ($\mathbf{X}\mathbf{X}^{\mathrm{T}}$). As PCA describes the data in terms of variance it is often necessary to normalise the data, typically in terms of standard deviations e.g. Auto scaling – Eq. (7).

In Fig. 4 there are four features, A, B, C and D. A and B are at 90°, cosine $90 = 0$, therefore they are uncorrelated. A and D are at 180°, cosine $180 = -1$, therefore they are anti-correlated. B and C are at 80°, cosine $80 = 0.1736$, therefore there is little correlation between them. C and D however are at 10°, cosine $10 = 0.9848$, they are therefore correlated.

PCA applications

The identification of bacteria that cause ear, nose and throat (ENT) diseases, *Staphylococcus* (*S.*) *aureus*, *Legionella pneumophila*, and *Escherichia* (*E.*) *coli* have

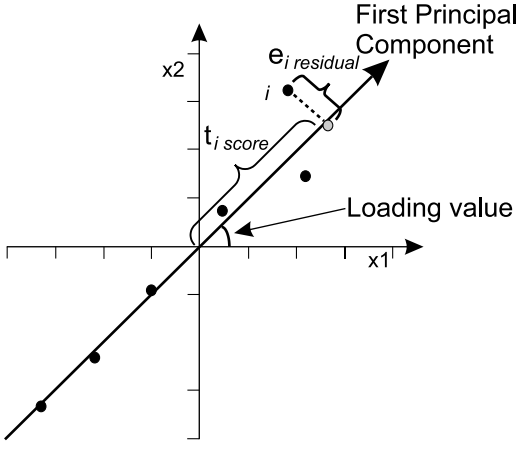

Fig. 3. PCA scores properties

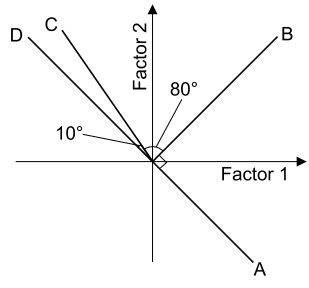

Fig. 4. PCA loading properties

been investigated and detected with a high degree of success [62]. It is a clinical advantage to be able to detect these pathogenic bacteria in the very early stages of their growth. A 32 CP array for the diagnosis of ENT infection [63] utilised PCA. Wet and dry swab samples were taken from patients and analysed by e-nose and standard microbiological techniques. Separate PCA scores plots of the wet and dry swab samples using the information from the standard microbiological tests showed visual clustering of different types of organisms. The organisms were found to give specific response patterns and thus were found to occupy distinct areas of the PCA scores plot. The application of PCA to e-nose data is facile to implement and the interpretation for this example involves the visual examination of separation between groups, this suitable for a non skilled operator to carry out. However if automation of the system were important, then a classification technique would be required.

Barko and Hlavay [64] characterised 4 GC stationary phases initially selected by differing McReynolds parameters as coatings on TSM sensors to 7 VOC offering diverse chemical properties using eigenvector information. It was found that the factors of the sensing compounds selectivity were dependent upon the VOC, and when the factors of the sensing compounds differed the VOC could be selectively distinguished. Guadarrama et al. [65] used a 6 sensor CP array to characterise two red wines from different geographical regions and one white wine. Two sampling techniques, static and dynamic headspace analysis were tested. PCA analysis showed that dynamic sampling out performed static by decreasing the cross interference of water and ethanol, thus permitting higher discrimination of the different wine types. The same team [66], used an array of 12 CP sensors on three white wines and two red. PCA analysis showed that sampling using a SPME fibre lead to noticeable differences in the responses of the sensors towards different wines and that the degree of selectivity obtained was not possible from ordinary sampling.

Partitional clustering algorithms

A cluster is a subset of the full data set; partitional clustering algorithms attempt to obtain partitions in the data instead of a clustering structure, such as the dendogram. A common problem with partitional methods is the need to specify the desired number of clusters before computation begins. The partitional techniques usually produce clusters by minimising a criterion function defined either locally or globally. Cluster prototypes are not usually known beforehand, and are calculated by the algorithm simultaneously with the partitioning of the data. Accordingly clustering techniques are among the unsupervised learning methods.

k-means clustering

The k-means algorithm works unsupervised and is popular as it is easily implemented. If n is the number of patterns, k is the number of clusters, and l is the number of iterations taken, the time complexity is O(nlk). Typically l and k are fixed so the algorithm has linear time complexity to the size of the data set. A major drawback to this algorithm is that it is sensitive to the selection of the initial partition and may converge to a local minimum of the criterion function if the initial partition is not well chosen [67].

A typical k-means algorithm minimises the squared error function:

$$E = \sum_{k=1}^{C} \sum_{i=1}^{n_j} \|\mathbf{x}_i^{(k)} - c_k\|^2 \qquad (6)$$

where C is the number of clusters, $\mathbf{x}_i^{(k)}$ is the i^{th} pattern belonging to the k^{th} cluster, and c_k is the centre of cluster k. The result is a hard partition of the data, assigning each pattern to a particular cluster. Several variants of the k-means algorithm have been reported [68], many attempt to estimate a good initial partition so that the algorithm is more likely to converge at the global minimum value. Another variation of the algorithm is to split and merge the resulting clusters. This type of variant allows an optimal partition starting from almost any arbitrary initial partition, providing well-determined threshold values are specified.

k-means applications

Grate et al. [69] used several exploratory methods including k-means, PCA and bar charts to show that 4 TSM sensors coated with different films of monolayer-protected nanoparticle (MPN) materials behaved differently to one another and poly(isobutylene) and poly(epichlorohydrin) sorptive polymer test cases.

Classifiers

A typical pattern classification system as summarised in Fig. 5, may be broken down into a number of independent stages incorporating:

1. Data acquisition system
2. Feature extraction system
3. Classifier
4. Decision making strategy

The conceptual boundary between a feature extractor and classifier proper is almost arbitrary. An ideal feature extractor would make a representation of data that reduces the classifiers job to trivial; conversely a universal classifier would not need a sophisticated feature extractor [70]. The task of a classifier is to use the feature vector provided by the feature extractor to assign the object it represents to a category. The degree of difficulty of the classification problem depends on the variability in the feature values for objects in the same category relative to the difference between feature values for objects in different categories. The variability in the feature values for objects in the same category may be due to complexity, and may be due to noise. The simplest measure of classifier performance is the classification error rate, the percentage of patterns that are assigned to the wrong category. The final stage of a pattern classification system is usually making a decision on the class assignment to the input patterns based on measurements taken from the selected features. E-noses are expected to be employed in particular environments and be small, portable, low power devices that are easy to operate. Given these qualities, Shaffer et al. [71] proposed 6 requirements that an ideal classification system should possess.

1. High accuracy – there should be as few misclassifications as possible.
2. Fast – for real-time analysis, the algorithm must be able to produce a classification with minimum delay.
3. Simple to train – in many applications the database of training patterns will be updated periodically and the classifier retrained. This procedure should be quick and simple to perform.
4. Low memory requirements – for small portable systems that may be used as handheld devices the classifier would need to consume few resources.
5. Robust to outliers – in uncontrolled environments the algorithm must be able to reduce the potential for misclassifications by being able to differentiate between a pattern on which it was trained to recognise and one that it was not. The assumption is that the system has been trained on all relevant patterns so any ambiguous patterns should be recognised as such.
6. Produce a measure of uncertainty – for many applications the algorithm needs to produce a measure of the match level of the classification, or a statistical measure concerning the certainty of the classification.

Testing and validation

Cross-validation a technique to compensate for an optimistic apparent error rate caused by training and testing on the same small data set. The apparent error rate is the percent of misclassified observations. The cross-validation routine omits each observation, one or more at a time (leave-one-out or leave-n-out) and recalculates the classification function using the remaining data, and then classifying the omitted observation. Splitting the data into two sets is used to calculate a more realistic error rate on larger data sets. One set to create the discriminant function and the other set to act as a validation set.

Discriminant function analysis

Discriminant function analysis (DFA) techniques are supervised and multivariate, they were first reported by Fisher [72, 73] on the well known iris data of Anderson [74]. There are the two types, linear and

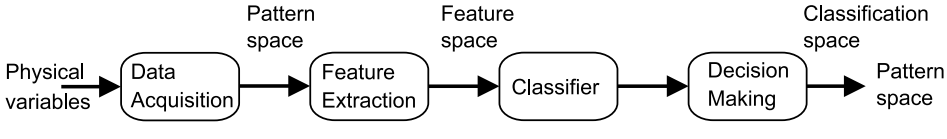

Fig. 5. Conceptual pattern classifier

quadratic, normal distribution is assumed for the classes in both cases [75].

Linear discriminant analysis

Linear discriminant analysis (LDA) finds a linear discriminant function (LDF) which is a linear combination of the of the original variables $\mathbf{x} = (x_1, x_2, \ldots, x_p)$, such that the ratio of the between-class scatter and the within-class scatter is maximized [45]. The class means are estimated $\mu_k = \frac{1}{N_k}\sum_{i=1}^{N_k}\mathbf{x}_i$ where N_k is the number of samples in class $k = 1, 2, \ldots K$ and N is the total number of samples. The true values of the distributions π are not known so are estimated, $\pi_k = \frac{N_k}{N}$. An assumption is made that covariance matrices are equal for all groups, so are pooled. An estimate of the pooled covariance matrix is:

$$\sum = \frac{1}{N-K}\sum_{k=1}^{K}\sum_{i=1}^{N_k}(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T \quad (7)$$

The linear discriminant function is:

$$\delta_k(x) = \mathbf{x}^T\sum^{-1}\mu_k - \frac{1}{2}\mu_k^T\sum^{-1}\mu_k + \log(\pi_k) \quad (8)$$

where $\sum^{-1}$ is the inverse of the covariance matrix, $\mathbf{x}^T$ is the transpose of the $\mathbf{x}$ matrix and μ_k^T is the transpose of the μ_k vector. The decision boundaries are linear equations in $\boldsymbol{x}$, and the boundary between two classes, A and B may be written:

$$\log\left(\frac{\pi_A}{\pi_B}\right) - \frac{1}{2}(\mu_A + \mu_B)^T\sum^{-1}(\mu_A - \mu_B) + \mathbf{x}^T\sum^{-1}(\mu_A - \mu_B) = 0 \quad (9)$$

where μ_A and μ_B are the class mean vectors for classes A and B, π_A and π_B are the estimated distributions of classes A and B. There are *K-1* discriminant functions available; these may be used for graphical display of the data.

Quadratic discriminant analysis

There is no assumption with quadratic discriminant analysis (QDA) that the groups have equal covariance matrices so they are estimated separately for each class; this requires more data than LDA for an accurate estimate.

$$\sum_k = \frac{1}{N_k}\sum_{i=1}^{N_k}(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T \quad (10)$$

The quadratic discriminant function is:

$$\delta_k(x) = -\frac{1}{2}\log\left|\sum_k\right| - \frac{1}{2}(\mathbf{x} - \mu_k)^T\sum_k^{-1}(\mathbf{x} - \mu_k) + \log(\pi_k) \quad (11)$$

Decision boundaries are quadratic equations in $\boldsymbol{x}$. QDA often fits the data better than LDA but has more parameters to estimate.

DFA applications

Di Natale et al. [76] used an 8 TSM sensor array in a preliminary study for the identification of lung cancer by breath analysis. Alkanes, benzene derivatives and aniline are possible volatile markers for lung cancer. Breath samples were taken from 35 volunteers with a form of lung cancer, 9 individuals after surgical therapy and 18 healthy volunteers. Analysis was carried out by partial least squares-discriminant analysis (PLS-DA), a linear technique. Correct classification of lung cancer patients was 100%, with 94% and 44% correct classification of non-cancer and post-surgical categories respectively.

Oliveros et al. [77] used a 12 MOS sensor nose to detect olive oil adulteration, including non-adulterated oil. Data was collected every second for 60 seconds then reduced to 8 points per sensor by selecting points at specific times; this was used as raw data. This was further reduced by PCA or other methods. LDA and QDA were applied to the extracted variables. BP-MLP networks were applied to the raw data set. The discriminant analysis techniques were found to give prediction rates above 95% for the oil samples. The BP-MLP gave worse results, probably due to the lack of feature selection.

Goschnick et al. [78] used a microarray of 38 MOS sensors for discrimination and recognition of tap water, ammonia contaminated water and chloroform contaminated water. Resistance values were normalised and then analysed by PCA and LDA. When sampling at 1 second response; using LDA with a leave-one-out cross validation, 93% of all points were correctly classified.

Gonzalez-Martin [79] employed a sensory array consisting of 6 MOS sensors with controlled temperature and humidity with LDA for the differentiation of Iberian breed swine hams according to their feeding

regimes. The classes investigated were feed only, feed plus acorns and acorns only. Initially 48 variables were to be used for the LDA (8 points per sensor), however due to software limitations a selection process was employed to choose only those giving differential information (actual number of variables used not given). The LDA gave results of 96% and 100% correct for raw and normalised readings respectively. The authors suggest that the e-nose with LDA can be considered a simple and rapid method for classification of Iberian breed swine on the basis of their diet.

k – nearest neighbour

The k – nearest neighbour algorithm (KNN) is a supervised hard partitional algorithm that is conceptually simple and easy to implement. Training the model simply consists of storing the training data. The distance between an unclassified pattern vector and every sample from the training set is calculated, it is assigned to the class of the k neighbours with the closest distances. The choice of the value of k is usually empirical, $k = 1$ is the most widely used and is known as the 1-NN or nearest-neighbour, the decision boundaries resulting are piecewise linear. 1-NN has an error rate which is twice as large as the Bayes optimum [80]. Larger values of k give smoother boundaries. The time complexity is $O(n^2)$ however, meaning that if the number of observations doubles the computational time quadruples. The distance metric most commonly used is the Euclidean.

KNN applications

Schiffman et al. [81] classified five fungal species commonly found in indoor environments using an array of 15 MOS sensors to sample air. LDA, KNN and least square analysis (LS) were employed to classify the data. The system was capable of discriminating among the fungi with up to 96% accuracy. It was also able to recognise five selected VOCs that are emitted by fungi (ethanol, 2-methyl-1-propanol, 3-pentanone, 3-octanol and 3-octanone). Carmel et al. [82] used a feature extraction method based on a curve fitting model of transient signals from arrays of either 8 MOS or 8 TSM sensors with KNN as the classifier. Classification rates of up to 97.8% were achieved.

Neural networks

Neural networks offer a powerful non-linear mapping capability, many types have been employed for e-nose data classification including Kohonen networks, learning vector quantization (LVQ) and its variations, MLP with variants of the backpropagation (BP) algorithm and ART.

Neural network architecture

A neural network is characterized by properties such as:

– the activation function used in the neurons (kernels)
– the network topology
– the learning (training) algorithm used

Multi-layer perceptrons

Of all the various artificial neural network (ANN) structures the multi-layer perceptron (MLP) is the most used. The MLP as shown in Fig. 6 are feed-forward networks with one or more layers of neurons between the input and output neurons. The additional layers contain hidden units that are connected to either the inputs or outputs by weighted connections. The capabilities of MLPs stem from the nonlinearities used within neurons.

Sigmoid function

The most popular continuous activation function used within MLP networks is the sigmoid or logistic function shown as Eq. , which is non-linear, scaled and differentiable. The sigmoid function kernel partitions the feature space with hyper-planes.

$$f(net) = \frac{1}{(1 + e^{-net})} \tag{12}$$

where net $= \sum input \times weight$, for a neuron.

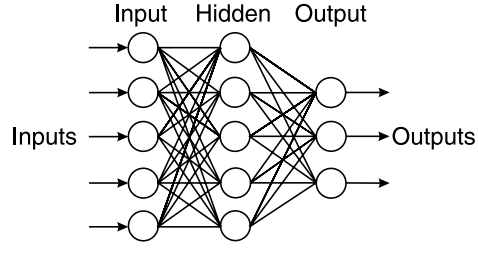

Fig. 6. A 3 layer MLP

Backpropagation training

The most used method for training the MLP is based on minimisation of a cost function and is known as the backpropagation (BP) algorithm [83]. Many alternate forms of backpropagation have been used; Jondarr [84] reviewed 65 varieties of this algorithm. The actual derivations for the formulae used in the BP algorithm come from the generalized delta rule. The delta rule is based on the error surface which represents cumulative error over a data set as a function of the network weights. Each possible network weight configuration is represented by a point on this error surface. A feed forward phase is first performed on an input pattern to calculate the net error. The partial derivative of the network error with respect to each weight gives information about the direction the error of the network is moving. If the negative of this derivative is taken (i.e. the rate change of the error as the value of the weight increases) and then added to the weight, the error will decrease until it reaches a local minimum. The error correction starts from the output layer to hidden layer weights, then to input layer weights, backwards through the network. Training continues until the error is minimised or a maximum number of training epochs has elapsed. A method often used to introduce adaptive learning into BP to lessen the possibility of chaotic behaviour is the momentum term.

The number of adjustable parameters P, within a MLP is dependent on the architecture used.

$$P = H + O + (I^{*}O) + (H^{*}O) \tag{13}$$

where I is the number of input neurons, H the number of hidden neurons and O the number of output neurons. Jurs et al. [31] recommended that there is at least twice the number of training patterns than adjustable parameters to properly adjust the network.

$$N_{train} \geq 2P \tag{14}$$

The number of hidden neurons is important for achieving generalization. In general, an increased number of neurons can result in over-training while a small number of neurons would normally result in under training (insufficient learning). The problem of finding an optimal setup is difficult as each specific configuration has a unique set of optimal parameters. These issues and others are discussed by Pardo [85]. The most commonly used analysis techniques for electronic nose data are PCA and MLPs. PCA is typically used for exploratory data analysis to see how the multivariate data is clustered and to assess the linear separability of the odour classes. MLPs are used to provide a predictive classification of unknown odour vectors from selected features.

BP-MLP applications

Garcia-Gonzalez and Aparico [86] used 18 MOS sensors to distinguish between lampante virgin olive oil from other categories. Temporal profiles were taken for each sensor and the data reduced by a windowed time slicing method. A wrapper genetic algorithm was applied to select the features for a neural network. Eleven features from 7 sensors were selected as input into a conjugate gradient trained MLP, producing 100% accuracy. This method is complex to apply but the benefits achieved in the classification accuracy are impressive; the results could be achieved with maximum response data only as the composition of the oils is unlikely to alter over the sampling time. This would help reduce the problem complexity as well as the processing time.

Pardo et al. [87] used a hierarchical system of PCA and BP-MLPs in cascade and a comparative SIMCA (Soft independent modelling of class analogy) technique to classify measurements for 14 olive oil types using a 12 MOS array. The PCA and BP-MLP was found to appreciably outperform the SIMCA for this task.

Zou et al. [88] used PCA as a filter to reduce 4 points from each of an array of 5 MOS sensors to 10 features. A BP-MLP was then used in a 10-8-2 configuration to distinguish between two types of vinegar with 98% accuracy, once again demonstrating the benefits of feature selection.

Radial basis function neural networks (RBF)

RBF networks are feed-forward connectionist architectures consisting of a hidden layer of kernels and an output layer of linear neurons. In RBF networks output units form a linear combination of the basis functions in the hidden layer. A basis function may be viewed as an activation function that produces a localised response to the input vector. The hidden processing elements are radially symmetric and have three properties:

1. A centre that is a vector in the input space, and which is typically stored in a weight vector from the input layer to the hidden processing element.
2. A distance measure to determine how far an input vector is from the centre.
3. A transfer function to evaluate the output of the hidden processing element by mapping the output of the distance function (e.g. Gaussian function).

There are several possible learning processes for a RBF network depending on how the centres of the hidden layer are specified. The different layers of a RBF network perform different tasks, so it is reasonable to separate the optimisation of the hidden and output layers by using different techniques [89]. The locations of the centres may be chosen randomly from the training data set or unsupervised clustering [90–92], or a genetic algorithm [93, 94] may be used to estimate the centres. An estimate of the variance of the input vector with respect to each centre provides the radius. Finally with the first two variables set, the weights may be calculated using a least mean squares algorithm, as per the BP algorithm. The most generalised form of supervised training is error correction learning [95] and allows simultaneous adjustment of the hidden neuron centres, the kernel width (spread parameter) and the output weights by a steepest descent method. RBF network decision boundaries are typically hyper-spheres.

RBF applications

Distante et al. [96] used an array of MOS sensors collecting data of several odorants such as pentanone, hexanal, water, acetone and three mixtures of these. A RBF network with 3 input and 6 output neurons was trained by starting with one hidden neuron and at each training iteration a new neuron was added until the network sum square error was below a threshold; then training stopped, resulting in a 3-157-6 configuration. The results were found to significantly outperform a BP-MLP, signifying that the adaptive insertion or deletion of hidden layer neurons in response to training errors is a powerful technique.

Szczurek and Maciejewska [97] used a 6 MOS based array to investigate the recognition of benzene, toluene and xylene in air with variable humidity. Response patterns were initially visualised using PCA and indicated that humidity was the main classification factor in the measurement dataset. LDA was able to classify benzene from toluene and xylene, however on occasion misclassification between toluene and xylene occurred. The first four PCs from the PCA were used as input into a four neuron input and one neuron output RBF network, this gave 100% classification. Demonstrating the power of feature selection and the non-linear decision boundaries of the RBF network.

Ali and co-workers [98, 99] classified different types of edible oil. Figure 7 shows the discriminant boundaries for four classifiers on this data, BP-MLP, RBF, QDA and LDA. The classes of edible oil show differing variance within groups and there is overlap

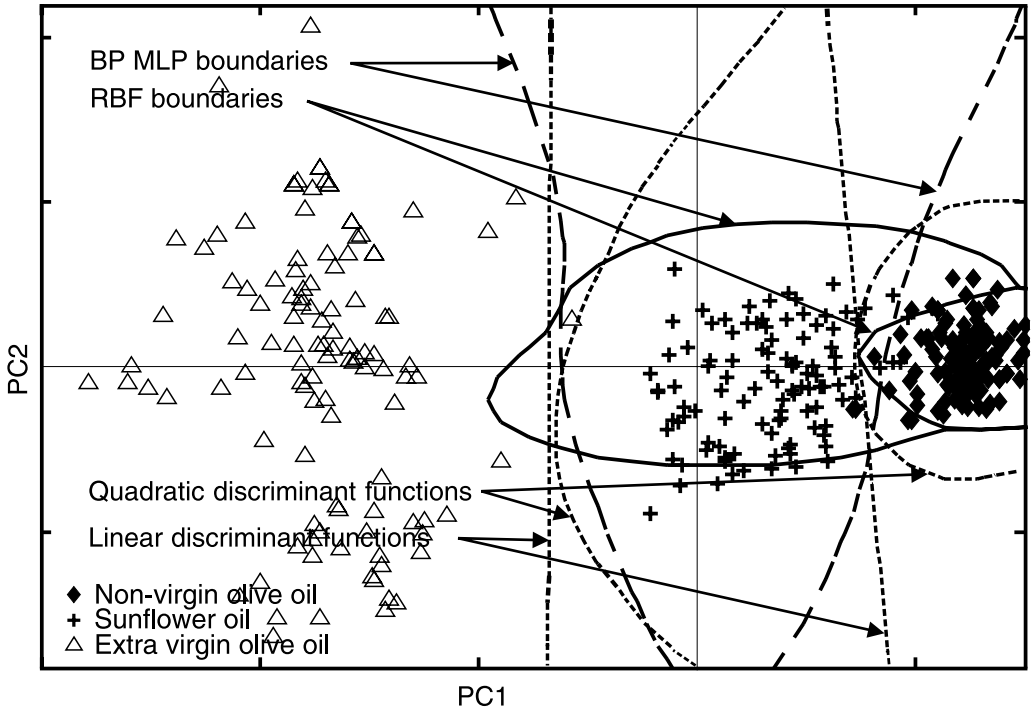

Fig. 7. Discriminant boundaries of BP-MLP, RBF, QDA and LDA, classifiers

between the non-virgin olive oil and sunflower oil. The DFA and ANN algorithms may yield comparable performance on well defined classification tasks, whereas ANNs usually obtain better performance on tasks that involve non-linear classification boundaries, such as where there is overlap. DFA techniques however have vastly reduced learning times.

Probabilistic neural networks

The probabilistic neural network (PNN) of Specht [100] is a parallel implementation of the Bayes statistical technique and is a four layer feed forward MLP. By replacing the sigmoid activation function often used in MLP networks with an exponential function (Gaussian), a PNN can compute nonlinear decision boundaries which approach the Bayes optimal. The decision requires an estimate of the probability density function (pdf) for each class. An estimated pdf is used in a Bayes decision rule with appropriate scaling, to produce a class estimate on the testing data. The PNN has only one adjustable parameter σ, the kernel width, which determines the degree of interpolation that occurs in determining the pdf. The PNN is limited to applications involving relatively small datasets; large datasets would lead to large network architectures and would increase the rate of misclassification.

PNN applications

McGill et al. [101] used a 3 SAW array for the detection of chemical agents and other toxic gasses and vapours. A PNN was able to distinguish between nerve, blister agents and over 20 different interferents with 100% accuracy. Lozano, Santos and Horrillo [102] used an array of 16 sputtered thin film MOS sensors to analyse white wines. The aromas of the wines were split into five categories; floral, fruity, herbaceous, microbiological and chemical. A PNN achieved 100% accuracy for floral, fruity, herbaceous and microbiological aromas; the accuracy to the chemical aromas was 97.2%. Santos et al. [103] used an array of 7 SAW sensors on Spanish wines of different grape varieties and ageing processes, a PNN classified 86% correctly.

Self-organising map

The Self-Organising Map (SOM), is a feed forward unsupervised learning network [104]. The SOM is a clustering type algorithm and is used for exploratory data analysis.

The SOM consists of a regular, typically two-dimensional grid of processing neurons (map units). Each unit is connected to adjacent ones by a neighbourhood relation. The number of map units used determines the accuracy and generalisation capability of the SOM. The map is trained iteratively; the best matching unit to the input is altered so that it closer matches that input, units close to the winner are also updated according to the neighbourhood relation. Figure 8 shows the architecture of a SOM and the linear relationship of neuron updating with the Euclidean distance from the winning unit. Although not a classification system SOM may be used as such by post processing the best matching map unit.

SOM applications

Davide [105, 106] used a SOM for a two-odour recognition problem and later extended the algorithm into a very large scale integration (VLSI) chip. The same group used the SOM system for the evaluation of tomato quality [107]. Tomatoes from conventional and organic land (classes) were split and labelled into 4 separate groups ranging from very good to poor due to their quality. PCA and SOM were then used to visually interpret the results. Although PCA showed a large separation for the poor organic tomatoes, the rest of the data was quite similarly distributed and it

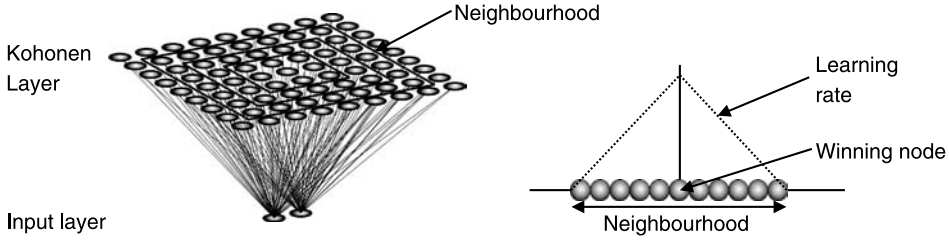

Fig. 8. Rectangular neuron arrangement and reducing neighbourhood scheme in SOM

was not possible to discriminate into groups. Using a SOM utilising 8 input neurons and a 10 by 10 map layer, it was possible to discriminate all four groups for each class of tomato and much less spread was observed within individual category groups.

Learning vector quantization

Learning Vector Quantization (LVQ) is a supervised variant of the SOM [108] and is a dedicated classifier. LVQ algorithms directly define class boundaries based on prototypes, a nearest neighbour rule and a winner-takes-all paradigm. The input space is covered with 'codebook vectors' (CV) each representing a region labelled with a class. A CV may be thought of as a prototype of a class member localised in the centre of a class decision boundary. A class may be represented by any number of CV, but one CV represents only one class. Class boundaries are built piecewise-linearly as segments of the mid-planes between CV of neighbouring classes. During learning the weights are modified, adjusting the position of a CV in the input space and therefore class boundaries. The winning CV is moved closer to the input, if it is labelled with the correct class, otherwise it is moved farther away. Classification after learning is on the label of the nearest CV to the sample.

LVQ applications

Kusumoputro [109] developed a neuro-fuzzy LVQ using a quartz resonator sensor system. Seyama et al. [110] used an array of 8 plasma polymerized film (PPF) coated TSM sensors on citrus and woody odours. LVQ was used to produce 'odour maps'. KNN processing resulted in 74% for the citrus and 91% accuracy for the wood odours respectively. Qian et al. [111] studied toxic vapour classification and concentration estimation for space suits after space walks. Studies were conducted using classifiers including LVQ, BP-MLP and PNN; results of up to 99% correct estimation of concentration of vapours such as hydrazine were obtained.

Fuzzy logic

Fuzzy set and system theory originated with the work of Zadeh [112, 113], and provides a superset to the traditional forms of logic and set membership that has predominated since the ancient Greeks. This theory extends the classical notion of true and false to include the range of real numbers [0, 1]. New operations for the calculus of logic were proposed which were generalisations of classic logic. In classical set theory an object either belongs to a set or it does not. Probability explains how events occur in random space; fuzzy logic includes situations where there is imprecision due to vagueness rather than randomness.

Fuzzy and crisp sets

A requirement of probability is that the mutually independent probabilities of a system must add to one, fuzzy membership functions do not possess this property. Fuzzy membership functions can be developed using a wide range of techniques including pdfs. Probability deals with the likelihood of an outcome and fuzzy logic deals with the degree of ambiguity. A probability of 1 indicates that the event is certain to occur. In fuzzy logic a membership of 1 means a complete lack of ambiguity. The linguistic statement 'there is a 50% possibility of a cloudy day' states the chance (0.5) of an ambiguous (cloudy) outcome. The membership function is the fundamental idea in fuzzy set theory. The values measure the degree to which objects satisfy imprecisely defined properties. A wide variety of membership functions can be used including triangular, trapezoidal and Gaussian.

Fuzzy c-means

Fuzzy c-means (FCM) is a clustering method of data analysis based on the fuzzy membership of each data point to each of the clusters of data formed [114]. The objective of the fuzzy c-means algorithm is to minimise the sum of the weighted squared distances between the data points, z_k and the cluster centres, v_i. The distances D_{ik}^2 are weighted with the membership values μ_{ik}. The objective function is then:

$$J(Z, U, V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D_{ik}^2 \qquad (15)$$

where:

$U = [\mu_{ik}]$ is the fuzzy partition matrix
$V = [v_1, v_2, \ldots, v_c]$ is a vector of cluster prototypes (centres)

$m \in (1, \infty)$ is a weighting exponent that determines the fuzziness of the resulting clusters, it is commonly chosen to be $m = 2$. As $m \to 1$, the partitioning becomes hard. Fuzzy clustering may be converted into

hard clustering by assigning each pattern to the cluster to which it has the highest membership.

D_{ik} may be determined by any appropriate norm, for example the Euclidean norm distance which results in point prototypes and spherical clusters.

Fuzzy c-means applications

Dutta et al. [115] used FCM, PCA and SOM to analyse data from a 32 CP array on eye bacteria and for the identification of *Staphylococcus aureus* infections in hospital environment [116]. They also used FCM and SOM feature extraction techniques along with a RBF neural network and achieved 100% correct classification for five different tea samples with different qualities from a 4 MOS array [117]. A commercial 6 sensor MOS based system was used by Marcelloni [118] to monitor the quality of doped food packaging material. Samples were taken every second for the duration of the sampling cycle of 166 seconds giving 696 points for each sample. Feature selection was performed by filtering using a supervised version of FCM (SFCM), giving an average of 47.6 features. These features were used in KNN for classification. It was found that using feature selection by SFCM increased accuracy from 79.58% to 82.15%.

Applications of other fuzzy systems

Dumitrescu et al. [119] used Fast Fourier Transforms (FFT) and PCA to filter features from an array of 8 CP sensors taking continuous samples in three phases. An unsupervised fuzzy divisive hierarchical classification system (UFDHC), basically a fuzzy binary search tree was then applied to some test cases including Hexanal, Nonanal and Hexanol, FFT was applied to the sensor signals keeping the first 3 coefficients – giving 24 features. PCA was applied to filter this to 9 features, the UFDHC was applied to these using a leave-one-out cross validation resulting in 100% correct classification. Lazzerini et al. [120] developed a fuzzy-based recogniser of olfactory signals (FROS) and applied it to samples taken from a 16 CP based system. Odorants acetone, butanol and o-xylene were combined into 7 mixtures. Each sample consisted of 450 readings for each of the 16 sensors. The output from a shape based and a dynamic based recogniser were combined to produce a list of fuzzy match pairs, odorant and match value, the highest match gave the classification – 82% of the samples were correctly classified. Singh, Hines and Gardner [121] compared the ability of a fuzzy neural network and a BP-MLP to classify odour samples that were obtained by an e-nose employing 4 or 12 MOS sensors. Coffee and tainted water samples were used as test cases. BP-MLP achieved best classifications of 86% on the coffee and 75% on the water data. The fuzzy network achieved 93% on the coffee and 85% on the water samples, Scott et al. [122] used a fuzzy set similarity classifier to discriminate between edible oils from an array of 6 TSM sensors. The data set of 346 samples was randomly split into 233 for training and 113 for testing. The technique was shown to be understandable and provided fast classification with discrimination rates up to 99% for the test set.

Adaptive resonance theory networks

Fuzzy ARTMAP (FAM)

Adaptive resonance theory was devised by Grossberg [123]. The aim was to solve some of the problems from which other neural networks suffer and to have a stable memory structure even with fast on-line learning that was capable of adapting to new data input, even forming totally new category distinctions. The most advanced model of the ART family is Fuzzy ARTMAP (FAM) [124, 125] which was developed for supervised slow learning. Unlike traditional MLP neural networks the architecture of FAM is self-organising; this phenomenon is known as the plasticity-stability dilemma; how a network retains learned patterns (stable) while remaining able to learn new ones (plastic). In a standard MLP network used for pattern classification an output neuron is assigned to every class of object that the network is expected to learn and must be trained off-line. In FAM the network dynamically assesses the assignment of output neurons to categories by competitive learning.

Two ART modules are interconnected by an associative memory and internal control structures as shown in Fig. 9. The ART_a module handles input patterns whilst ART_b the class patterns, both modules use a competitive strategy in updating, fuzzy comparisons are made between the input pattern and the neuron weights. The associative memory maps ART_a patterns to ART_b classes. The orienting subsystem is responsible for generating a reset signal whilst the gain control sums the input signal. The network is able to perform real time learning without losing previously learnt patterns by using an incremental weight update procedure known as slow recoding.

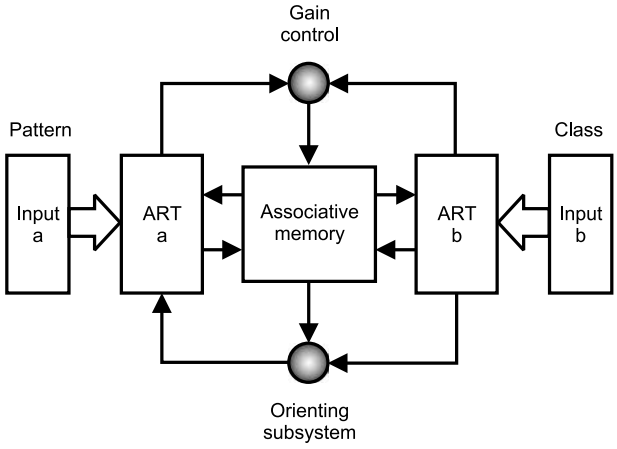

Fig. 9. FAM architecture

Simplified fuzzy ARTMAP (SFAM)

The main drawback to the ART family of networks is the complicated architectures instead of simple algorithms. This problem has been tackled by several authors such as Kasuba [126] who developed SFAM which is a simplified version of FAM. SFAM aims to remove the redundancies in FAM and so trains much faster. Vakil-Baghmisheh and Pavesic [127] report that their version of SFAM is capable of obtaining almost Bayes optimal classification rates with an increased training speed of over 50% compared to that of Kasubas' version.

Applications of FAM

Llobet et al. [128] used FAM on cows breath data from a 6 MOS array. The samples were split between healthy, ketotic and sub-clinical ketotic. The network performed with an accuracy of 79%, higher than the accuracy from a comparative BP-MLP, whilst also providing quicker training. Vinaixa et al. [129] used a MS e-nose with a solid-phase microextraction (SPME) preconcentration system to collect data of fungal growth in bakery products. Averages of mass spectra along detected peaks were used as sensor measurements. PCA was applied to the response matrix, and the first 10 PCs used as data for further processing. DFA was performed and the range scaled Eigenvalues obtained used as input in FAM. Predictions of the correct fungal genus reached 78% and 88% after 24 and 96 hours of incubation respectively. The same group used the MS e-nose to detect the rancidity of potato crisps [130]. FAM classification using 78 input variables was used, the average classification rates of crisp samples in four rancidity stages was 88%. A PCA filter was used to reduce the number of features to 10; with this input the classification of crisp rancidity improved to 93%. This shows although the algorithm can train rapidly on large data sets, redundancy within the data is still an important condition that must be addressed for optimal performance. Selecting more than 10 PCs did not improve the results. Ali [131] used SFAM on data from 6 TSM sensors to discriminate between edible oil samples with over 99% accuracy. The algorithm was found to offer easy implementation to maximum sensor response data from the e-nose whilst rapidly providing automation capable results.

Growing networks

The ART family and derivatives are not the only adaptive networks, the growing cell structure (GCS) network of Fritzke [132] is based on the SOM, but comprised of k-dimensional simplices. The value of k is typically 2, making the simplices triangular, a new node is inserted every λ iterations, where λ is a constant. The new node is positioned to support the node that has accumulated the highest error during the previous cycle. The network grows and adapts until a stopping criteria is met, either a predefined size or the error has reached an acceptable level. The growing neural gas (GNG) network, also of Fritzke [133] adds nodes every λ iterations but the structure is not constrained. The grow when required (GWR) network of Marsland et al. [134] may add a new node at any time, positioned dependent on the input and current winning node. Cheng and Zell [135] used the maximum response data from an array of 12 TSM sensors of seven coffee brands and achieved 86.7% correct classification from a GCS network.

Comparisons of classification techniques

Shaffer et al. [71] ran a comparison of 7 pattern recognition systems including LVQ, PNN, BP-MLP, KNN and LDA from simulated and actual data from 4 and 6 SAW sensor sets for chemical warfare agents. Six criteria were examined, speed of operation, simplicity of use, memory requirements, robust handling of outliers, a statistical measure of uncertainty and classification rate. BP-MLP achieved 93% accuracy, was fast but not simple to train and had low memory requirements. KNN gave 91% accuracy, was slow to operate

but was simple with high memory requirements. PNN gave 94% accuracy was slow but simple to operate but had high memory requirements. LVQ gave 95% accuracy was fast, sometimes simple to train and had low memory requirements. PNN was recommended for applications where a confidence measure and fast training are critical, but speed of classification and memory requirements are not, and LVQ for all other applications.

The performance of LVQ, BP-MLP and FAM were compared in determining the ripeness of bananas using a MOS sensor based system by Llobet et al. [136]. LVQ correctly classified 92% of banana samples into seven stages of ripeness, whilst FAM and MLP networks performed 90% and 83%, respectively. FAM however was shown to learn new samples without degrading in performance on previous samples and trained more quickly than both BP-MLP and LVQ. Gardner et al. [62] compared the performance of LVQ, BP-MLP and FAM whilst classifying potable water containing cyanobacteria using a 6 MOS sensor array. It was found that LVQ gave a slightly better prediction rate of 95.1% than the other two methods which both gave 92.3%, but that FAM used an order of magnitude less iterations to train.

Vaid et al. [137] compared several discriminant algorithms including, KNN, LDA and QDA on data from an array of 20 carbon black CP sensors using data from compositionally similar pairs of compounds. KNN performed the worst. LDA was found to perform the best, better than QDA, a surprising result attributed to the covariance matrix estimates of QDA being based on a class by class basis rather LDA using pooled data, so being less likely to reflect the true covariance matrices; the use of more data could have solved this problem.

Martin et al. [75] compared pattern recognition techniques on data taken by an array of 6 MOS sensors on vegetable oils. Data was taken every 1 second for 60 seconds for each sensor. The feature numbers were reduced by selecting 8 points on the curve for each sensor, resulting in 48 points for each sample. KNN, QDA and BP-MLP were applied to this feature set and a LDA reduced data set consisting of 9–36 points. KNN produced 80%, QDA and BP-MLP up to 94% correct. The better classification rates were achieved using lower feature numbers.

An array of 8 carbon black CP sensors was used by Bicego et al. [138] to discriminate between acetone, ethanol and 2-propanol using KNN and a MLP. The MLP was trained using a reactive Tabu search. Using all 8 features KNN gave a classification rate of 99%, whilst the MLP gave 100%. Using filter reduction by PCA to 3 features both classifiers achieved 100% correct predictions.

Dutta et al. [139] used an array of 4 MOS sensors to collect data for the determination of egg freshness. Measurements were taken every 5 seconds over the sampling period but only the maximum response was used in analysis. The data was normalised by dividing each reading by the maximum value. PCA, SOM and FCM were applied to establish data cluster and that results were not arbitrary; they all corresponded with the three egg freshness states. BP-MLP, LVQ, PNN and RBF networks were applied as classification systems. A BP-MLP using a 4-4-3 configuration gave 71%, LVQ with 4-3-3 gave 84%, PNN gave 89% and RBF 90% correct classifications. Training times were 7 hours for MLP, LVQ took 2 hours, RBF took 61 minutes and PNN took 57 minutes. It was concluded that LVQ was a good solution, but RBF preferred.

Kuske et al. [140] used a MOS sensor array to detect *Aspergillus versicolor* growing on different building materials. Moulds are known to produce a wide range of VOCs, including alcohols and ketones, which make their detection by e-nose feasible. The standard method for their detection would involve collection onto an adsorbent which would then be used in GC/MS analysis. A comparison of KNN and fuzzy KNN classifiers was studied on the mould sample data. It was found that both algorithms performed with very similar classification rates with the KNN slightly outperforming the fuzzy KNN. Using LDA as a filter technique the fuzzy KNN gave the highest classification rate of 89%.

Kinetics modelling

Some research groups have used part of or the entire response curve of the sensors, either during the sampling phase, or the whole sample/purge cycle to help improve separation between analytes. The shape of the response curve varies according to the analyte-sensor interactions. Saunders et al. [141] used normalised kinetic profiles of TSM sensors in conjunction with an BP-MLP as a classification technique. They found it was possible to identify different compounds by the change in shape of the response curve observed in the normalised results. Freeman et al. [142] using TSM sensors, suggested that by building a database of

response curves for compounds, faster classification would be possible by curve fitting of the initial response readings. Wavelet transforms have been applied as a feature extraction and data compression technique (mapping features) to temporal profiles [143–146]. Total time profiles were analysed by Skov and Bro [147] by using PARAFAC2 multi-way modelling. PARAFAC is an extension of PCA producing tri-linear models of the underlying data. PARAFAC2 is a development of the original PARAFAC which aims to handle generally varying profiles in an efficient manner, instead of one loading matrix being produced there is a time loading matrix for each sensor. Tests were made on data taken from liquorice samples using 12 MOS sensors. There were three groups, good, bad and fabricated bad. Classifications were made by KNN using the Euclidean distance of scores and compared to PCA processed data. PAFAFAC2 processed data produced marginally better results than PCA, 59 correct from 60 samples over 57 correct for PCA.

Sensor selection

There are several factors that need to be addressed for selection of sensors, including drift, robustness and reliability. Feature selection algorithms have three goals, to improve classification accuracy, to reduce the cost of extracting features and to improve the reliability of the performance. Large feature sets often contain redundant information that degrade the accuracy and speed of the classifier. Initial sensor selection may be achieved by a rational analysis of the vapour-sorbent interactions of each sensor. This involves careful analysis of the components of the test volatiles to cover what Grate [148] calls ''coating space''. It is pointless to select a set of sensors from a pool that does not include the most suitable for the task.

Wilson et al. [149] used a 30 MOS array with rank-ordering feature extraction on samples of breath alcohol mixtures (beer, wine, vodka). The centroid of each cluster was computed, then the Euclidean distance to each of the samples. These distances were then ranked according to the closeness to a particular centroid. The samples closest to each centroid were filtered out by determining whether they belonged to the cluster. From this, a centroid proximity metric was calculated and a ranking of each sensor contribution to this produced. The results from the full feature set were compared to the reduced set using BP-MLP and RBF networks. The feature extraction improved the classification rate for the BP-MLP from 92% to 95% and the RBF from 87% to 94%.

Carey et al. [150] analysed the results from 27 TSM GC stationary phase coated sensors to discriminate between mixtures of up to 14 analytes. PCA eigenvectors were analysed and using the contribution of each sensor to the highest ranking eigenvectors a subset of 7 sensors were selected that provided the maximum variance within the data. Lau et al. [151] considered the linear solvation energy relationship (LSER) to select several GC stationary phase coated TSM sensors. The results of testing were examined by the PCA eigenvector technique and a subset of 8 sensors was derived. Kermani et al. [152] used PCA eigenvector analysis on an array of 32 CP sensors on coffee data to select a subset of only 3, these gave no degradation in classification performance. This technique selects features that produce the highest variation, but does not guarantee that the selected features give the best classification rate as it ignores the contribution a feature makes to classification.

An alternate approach to feature selection is to search for combinations of sensors that produce the highest classification rate. Park et al. [153] used the responses to mixtures of subsets of 16 organic vapours collected by 6 polymer coated SAW sensors. Monte-Carlo simulations coupled with pattern recognition analysis was used to derive statistical estimates of vapour recognition rates as a function of the number of sensors in the array. It was found up to 4 vapour subsets could be recognised by an equal number of sensors to the number of vapours. It was also found that for 6 vapour subsets increasing the number of sensors did not improve performance. They concluded that large arrays are not needed for accurate vapour recognition and quantification. Benedetti et al. [154] analysed the data from 70 honey samples from a 12 MOS and 10 MOSFET sensor array. Using bar charts of sensitivity for the sensors as a filter, it was found that only three sensors responded well to the samples. A BP-MLP using 3 inputs was able to separate the honey into their four regions of origin. Pardo et al. [155] selected sensors from a hybrid array of 7 TSM, 8 MOS and 4 EC sensors. An exhaustive search using cross-validated KNN ($k = 3$) on subsets of 2, 3 and 4 was performed from a pool of 19 candidate sensors. PCA scores plots were used to show an improvement in discrimination from all sensors to the best performing two sensor subset.

The sequential forwards and sequential backwards selection (SFS and SBS) algorithms are commonly used search methods for feature selection. They apply a local search to test the current feature subset. The algorithms begin with a single solution and iteratively add or remove a feature if it provides the best improvement in performance, until some termination criterion is met. The forward type starts with the empty set and adds features, the backward starts with the full set and removes features. They cannot backtrack however; that is, to reinstate a feature that has been discarded.

Eklov et al. [156] selected features using the sequential forwards search (SFS) method from the transient signals of an array of 7 pt-MOSFET sensors monitoring *E. coli* biomass concentration. The root mean square error (RMSE) from a multi-linear regression model was used as the selection criterion. It was found that 10 features from a set of 85 gave the best prediction performance. The SFS algorithm was used as a wrapper around a BP-MLP by Polikar et al. [157] to select a subset of sensors from an array of 12 TSM sensors when classifying 12 VOCs. It was found that 2–4 sensors could adequately classify a VOC, and that the performance of classifiers typically degrades as the number of sensors increase beyond that number. It was noted that the SFS method is prone to being trapped in local maxima in the performance space.

Genetic algorithms

A genetic algorithm (GA) operates on a population of potential solutions applying the principle of survival of the fittest to produce successively better approximations to a solution. The population of potential solutions allows the technique to be massively parallel in operation [158]. At each generation of a GA, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and reproducing them using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals from which they were created; mimicking natural adaptation. A fitness or objective function is a method of calculating how good or bad the individual solutions within the population are at solving the specific problem. A GA may be used as a wrapper technique for a classifier; the GA selects the features and possibly the classifier configuration.

GA applications

Schiffman et al. [159] used data from an array of 32 CP sensors on a medication spray off-odour. Time response curves of 30 points for each sensor were reduced to mapping features using windowing functions and integrating to 4 points per sensor, and then further reduced by PCA to one point per sensor. PCA scores of different time slices suggested that the Freon propellant of the spray was carrying leached compounds from the container. Kermani et al. [160] used the same complicated filtering technique on fragrance data. A LM-MLP (Levenberg-Marquardt trained MLP) was set heuristically and results compared to a GA wrapped LM-MLP. The GA selected feature input and set parameters for the LM-MLP. Classification rates were 86.87% for the LM-MLP and 95.23% for the GA wrapped LM-MLP, demonstrating that the GA controlled feature and network parameter selection was a valuable aid in improving classification rates.

A comparison study of some neural network techniques conducted by Jurs et al. [31] used data taken by a polymer-coated fiber optic sensor array on 20 organic vapours. Feature selection was performed using a GA wrapper with a LVQ classifier and achieved 90% correct classification; BP-MLP networks set heuristically only performed 73% correct classification. A 14 CP sensor based system using a dynamic bubbler for sampling was used by Pavlou et al. [161] to detect and discriminate between 6 gastroesophageal isolates, including *Staphylococcus aureus* and *Helicobacter pylori* from the headspace of complex broth cultures. Four points were taken from each sensor giving 56 features per sample. A BP-MLP gave a prediction rate of 98%. Data from the headspace of *H. pylori* on three different media (enriched, normal and sterile) was processed by a GA wrapped BP-MLP. A 16 feature subset was found to give a correct prediction rate of 94%. Corcoran [162, 163] used a GA wrapper technique around a distance measure based on normalised mean distances between pairs of classes to reduce a feature set of 208 to 21, simultaneously increasing classification rates from 69% to 93%.

Data fusion from 4 different electronic noses was used to classify fruit solutions by Boilot et al. [164]. Range scale normalisation was performed on the data sets so that they could be combined. The fused data

contained 72 features and gave a prediction rate of 86.7% using RBF and 80% using PNN. Selecting a subset of 8 features by analysing PCA loadings produced the same classification rates. However using a GA wrapped PNN classifier to select 6 features resulted in 95% correct predictions.

Aleixandre et al. [165] conducted tests on mixtures of air pollutants such as NO_2, toluene and octane using an array of 8 MOS sensors. Dynamic sampling was used and several features taken from the transient signal. BP-MLP and PNN were used in classification tests on different feature subsets. It was found that as more features were added to the input the networks took longer to train and the results barely improved. When using all of the features the BP-MLP was no longer able to converge. A GA with a PNN fitness function was used to extract feature subsets, it was found that 6 features performed as well as the full feature set. The BP-MLP gave a slightly higher classification rate using this feature set than the PNN, but took longer to train.

Llobet and co-workers [166] used a GA wrapper for a FAM classifier to analyze single vapours and binary mixtures of three VOC. Data was taken by a 12 element MOS array and 10 samples used from the dynamic curves. The GA allowed the number of features used to be reduced from 120 to 9, which significantly increased the generalization of the FAM classifier. Classification rates of 91.67% and 88.33% were reached for single vapours, and single vapours and their binary mixtures respectively. Gardner et al. [167] used a fixed length integer gene based GA wrapper around a PNN to select subsets of specific numbers of features from 32 carbon black CP sensors and found that 5–6 sensors could achieve almost as high classification rates as 32. The fixed length of the gene implies that the optimum number of sensors is known before the selection procedure is applied.

Wrapper type search methods have been found to be useful for feature selection, although they do not prevent irrelevant or correlated features from being selected. For best classifier performance the features selected would also need to be checked for correlation to ensure that redundant or irrelevant information did not exist. To address this Kermani et al. [168] used their windowing functions and integration before filtering by PCA to produce uncorrelated features, and then a GA wrapper around a LM-MLP to select a combination of features to provide the best classification performance. This very complicated pre-processing and feature selection process will produce features that are both uncorrelated and provide the best separation for classification; it does not, however, help to reduce the cost of extracting features. A simplification and refinement of this technique may provide a generalised, practical method of feature selection to ensure that a minimum number of features are both uncorrelated and provide a maximum classification rate.

Drift calibration and counteraction

The stability of a sensor is its ability to give the same output when measuring a constant input, measured over a period of time. The term *drift* is used to describe the change that may occur. Some sensors types are prone to drift. One of the critical limitations of e-nose systems is measurement scattering over time due to sensor drift, this causes data clusters to shift or merge over time gradually reducing the systems classification performance. If this drift is predictable it may be compensated for, or a counteraction scheme put into place [169]. The alternative is frequent recalibration of the system to ensure that the prediction model remains valid, such as using a reference analyte [170]. Sisk and Lewis [171] found that a simple linear sensor by sensor calibration scheme was effective at restoring classification performance in binary separation tasks, the frequency of which was determined by occasional testing of the LDA classifier. Adaptive models using SOMs [172, 173] use newly collected data that match stored analyte signatures to continuously retrain the classifier. This technique is simple as no actual recalibration is performed by the user. Llobet et al. [136] and Gardner et al. [174] found that FAM was plastic enough to learn new samples without degrading in performance on previous samples. The drift, however, must be gradual as a discontinuity in response would immediately invalidate the classification model. This method also assumes that the sensors are drifting and not the analyte composition/concentration. The calibration and standardisation of drift counteraction algorithms is an ongoing topic of research by the NOSE II network [175].

Conclusions

A chemical sensor array system may detect and identify any analyte, provided certain criteria are met. The

array must contain some sensors that are sensitive to the analytes of interest, and these must be selective enough to distinguish between them and all others. These sensors must be stable enough to give consistent responses during the training phase and later use. The array should be trained on all potential analytes that may be encountered, and employ a pattern classification algorithm that is suitable for the sensor response patterns that will be generated. For example, some classifiers work well for binary separations of data that is largely multivariate Gaussian (e.g. LDA), while others work well for classifying many groups of highly nonlinear data (e.g. ANN).

The above criteria reduce the flexibility of any chemical sensor array, particularly for situations for which it is not known which analytes that may be encountered. Without this knowledge, it is not feasible to tailor an array with a suitable number of task-appropriate sensors, and impossible to train the array on every possible analyte. Even if it were possible to do so, retraining would be occasionally necessary if there was drift in the sensor responses. Choosing an inappropriate classification algorithm could result in poor results, either because the algorithm lacks sufficient plasticity to model nonlinear data, or because a highly plastic model was over trained on an insufficient number of training data. This situation has restricted applications to situations in which it is known approximately what to search for. Rarely have applications been applied where it is not known what analyte may be encountered because of the sensor selection and training required.

In the past simple analysis techniques were used to classify patterns, increasingly there has been a shift towards powerful mapping systems such as MLP, RBF, LVQ, PNN and FAM. MLPs are often outperformed by LVQ and FAM. These techniques and other self adjusting methods are becoming the benchmarks for predictive classifiers in VOC recognition. Neuro-fuzzy methods are being increasingly used as they have the ability to cope with imprecise data in a predictable way such as noise and drift. Continuous measurements from large array based systems produce a data rich environment; data processing becomes problematic. It is a challenge to be able to reduce the dimensionality of the incoming data whilst preserving the relevant information. Analysis techniques are rarely used alone, raw data is feature extracted, plotted and classified. Sophisticated sensor selection to target specific the differences between analytes are starting to become used, this should lead to more application areas.

References

[1] Persaud K, Dodd G (1982) Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. Nature 299: 352

[2] Albert K J, Lewis N S, Schauer C L, Sotzing G A, Stitzel S E, Vaid T P, Walt D R (2000) Cross-reactive chemical sensor arrays. Chem Rev 100: 2595

[3] James D, Scott S M, Ali Z, O'Hare W T (2005) Chemical sensors for electronic nose systems. Microchim Acta 149: 1

[4] Janata J, Josowicz M, DeVaney D M (1994) Chemical sensors. Anal Chem 66: 207R

[5] Strike D J, Meijerink M G H, Koudelka-Hep M (1999) Electronic noses – A mini-review. Fresenius J Anal Chem 364: 499

[6] Snopok B A, Kruglenko I V (2002) Multisensor systems for chemical analysis: state-of-the-art in electronic nose technology and new trends in machine olfaction. Thin Solid Films 418: 21

[7] Janata J, Josowicz M, Vanysek P, DeVaney D M (1998) Chemical sensors. Anal Chem 70: 179R

[8] Janata J (2001) Centennial retrospective on chemical sensors. Anal Chem 73: 150A

[9] Nagle H T, Gutierrez-Osuna R, Schiffman S S (1998) The how and why of electronic noses. IEEE Spectr 35: 22

[10] Gardner J W, Bartlett P N (1999) Electronic noses: principles and applications. Oxford University Press, Oxford

[11] Pearce T C, Schiffman S S, Nagle H T, Gardner J W (2003) Handbook of machine olfaction-electronic nose technology. Wiley-VCH, Weinheim, p 1265

[12] Peres C, Begnaud F, Eveleigh L, Berdague J-L (2003) Fast characterization of foodstuff by headspace mass spectrometry (HS-MS). TrAC Trends Anal Chem 22: 858

[13] Marti M P, Busto O, Guasch J, Boque R (2005) Electronic noses in the quality control of alcoholic beverages. Trac-Trends Anal Chem 24: 57

[14] Mielle P (1996) 'Electronic noses': Towards the objective instrumental characterization of food aroma. Trends Food Sci Technol 7: 432

[15] Bartlett P N, Elliot J M, Gardner J W (1997) Electronic noses and their application in the food industry. Food Technol-Chicago 51: 44

[16] Schaller E, Bosset J O, Escher F (1998) 'Electronic noses' and their application to food*1. Lebensmittel-Wissenschaft und-Technologie 31: 305

[17] Garcia-Gonzalez D L, Aparicio R (2002) Sensors: From biosensors to the electronic nose. Grasas Y Aceites 53: 96

[18] Ampuero S, Bosset J O (2003) The electronic nose applied to dairy products: a review. Sens Actuators B 94: 1

[19] Deisingh A K, Stone D C, Thompson M (2004) Applications of electronic noses and tongues in food analysis. Int J Food Sci Technol 39: 587

[20] Gardner J W, Shin H W, Hines E L (2000) An electronic nose system to diagnose illness. Sens Actuators B 70: 19

[21] Thaler E R, Kennedy D W, Hanson C W (2001) Medical applications of electronic nose technology: review of current status. Am J Rhinol 15: 291

[22] Gostelow P, Parsons S A, Stuetz R M (2001) Odour measurements for sewage treatment works. Wat Res 35: 579

[23] Nimmermark S (2001) Use of electronic noses for detection of odour from animal production facilities: a review. Wat Sci Tech 44: 33
[24] Bourgeois W, Burgess J E, Stuetz R M (2001) On-line monitoring of wastewater quality: a review. J Chem Technol Biotechnol 76: 337
[25] Bourgeois W, Romain A C, Nicolas J, Stuetz R M (2003) The use of sensor arrays for environmental monitoring: interests and limitations. J Environ Monitor 5: 852
[26] Kuske M, Romain A-C, Nicolas J (2005) Microbial volatile organic compounds as indicators of fungi. Can an electronic nose detect fungi in indoor environments? Build Environ 40: 824
[27] Liden H, Mandenius C-F, Gorton L, Meinander N Q, Lundstrom I, Winquist F (1998) On-line monitoring of a cultivation using an electronic nose. Anal Chim Acta 361: 223
[28] Bachinger T, Riese U, Eriksson R, Mandenius C F (2000) Monitoring cellular state transitions in a production-scale CHO-cell process using an electronic nose. J Biotechnol 76: 61
[29] Liden H, Bachinger T, Gorton L, Mandenius C F (2000) On-line determination of non-volatile or low-concentration metabolites in a yeast cultivation using an electronic nose. Analyst 125: 1123
[30] Deisingh A K, Thompson M (2004) Biosensors for the detection of bacteria. Can J Microbiol 50: 69
[31] Jurs P C, Bakken G A, McClelland H E (2000) Computational methods for the analysis of chemical sensor array data from volatile analytes. Chem Rev 100: 2649
[32] Bellman R (1961) Adaptive control processes: a guided tour. Princeton University Press, Princeton, N.J.
[33] Jain A K, Duin R P W, Mao J C (2000) Statistical pattern recognition: a review. IEEE T Pattern Anal Machine Intelligence 22: 4
[34] Haykin S (1999) Neural networks, a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River, N.J., p 842
[35] Blum A L, Langley P (1997) Selection of relevant features and examples in machine learning. Artificial Intelligence 97: 245
[36] Hyvärinen A, Karhunen J, Oja E (2001) Independent component analysis. John Wiley & Sons, New York, p 504
[37] Stone J V (2002) Independent component analysis: an introduction. Trends Cogn Sci 6: 59
[38] Kohavi R, John G H (1997) Wrappers for feature subset selection. Artif Intell 97: 273
[39] Mertens B, Thompson M, Fearn T (1994) Principal component outlier detection and simca – a synthesis. Analyst 119: 2777
[40] Walmsley A D, Haswell S J, Metcalfe E (1991) Methodology for the selection of suitable sensors for incorporation into a gas sensor array. Anal Chim Acta 242: 31
[41] McAlernon P, Slater J M, Lowthian P, Appleton M (1996) Interpreting signals from an array of non-specific piezoelectric chemical sensors. Analyst 121: 743
[42] Park J, Zellers E T (2000) Temperature and humidity compensation in the determination of solvent vapours with a microsensor system. Analyst 125: 1775
[43] Groves W A, Zellers E T, Frye G C (1998) Analyzing organic vapors in exhaled breath using a surface acoustic wave sensor array with preconcentration: selection and characterization of the preconcentrator adsorbent. Anal Chim Acta 371: 131
[44] Groves W A, Zellers E T (2001) Analysis of solvent vapors in breath and ambient air with a surface acoustic wave sensor array. The Annals of Occupational Hygiene 45: 609
[45] Johnson R A, Wichern D W (1982) Applied multivariate statistical analysis. Prentice-Hall International, Englewood Cliffs, New Jersey
[46] Brezmes J, Ferreras B, Llobet E, Vilanova X, Correig X (1997) Neural network based electronic nose for the classification of aromatic species. Anal Chim Acta 348: 503
[47] Jonsdottir R, Olafsdottir G, Martinsdottir E, Stefansson G (2004) Flavor characterization of ripened cod roe by gas chromatography, sensory analysis, and electronic nose. J Agric Food Chem 52: 6250
[48] Gan H L, Tan C P, Che Man Y B, NorAini I, Nazimah S A H (2005) Monitoring the storage stability of RBD palm olein using the electronic nose. Food Chem 89: 271
[49] Gan H L, Man Y B C, Tan C P, NorAini I, Nazimah S A H (2005) Characterisation of vegetable oils by surface acoustic wave sensing electronic nose. Food Chem 89: 507
[50] Everitt B (1974) Cluster analysis. Heinemann Educational, London
[51] Sammon J W (1969) A nonlinear mapping for data structure analysis. IEEE T Comput c-18: 401
[52] Lerner B, Guterman H, Aladjem M, Dinsteint I, Romem Y (1998) On pattern classification with Sammon's nonlinear mapping an experimental study. Pattern Recogn 31: 371
[53] Pal N R (1999) Soft computing for feature analysis. Fuzzy Sets Syst 103: 201
[54] Kowalski B R, Bender C F (1972) Pattern recognition. Powerful approach to interpreting chemical data. J Am Chem Soc 94: 5632
[55] Kowalski B R, Bender C F (1973) Pattern recognition. II. Linear and nonlinear methods for displaying chemical data. J Am Chem Soc 95: 686
[56] Persaud K C, Khaffaf S M, Payne J S, Pisanelli A M, Lee D H, Byun H G (1996) Sensor array techniques for mimicking the mammalian olfactory system. Sens Actuators B 36: 267
[57] Byun H G, Persaud K C, Khaffaf S M, Hobbs P J, Misselbrook T H (1997) Application of unsupervised clustering methods to the assessment of malodour in agriculture using an array of conducting polymer odour sensors. Comput Electron Agr 17: 233
[58] Persaud K C, Pisanelli A M, Szyszko S, Reichl M, Horner G, Rakow W, Keding H J, Wessels H (1999) A smart gas sensor for monitoring environmental changes in closed systems: results from the MIR space station. Sens Actuators B 55: 118
[59] Bourgeois W, Hogben P, Pike A, Stuetz R M (2003) Development of a sensor array based measurement system for continuous monitoring of water and wastewater. Sens Actuators B 88: 312
[60] Biswas S, Heindselmen K, Wohltjen H, Staff C (2004) Differentiation of vegetable oils and determination of sunflower oil oxidation using a surface acoustic wave sensing device. Food Control 15: 19
[61] Nonaka A, Tanaka M, Anguri H, Nagata H, Kita J, Shizukuishi S (2005) Clinical assessment of oral malodor intensity expressed as absolute value using an electronic nose. Oral Diseases 11: 35
[62] Gardner J W, Shin H W, Hines E L, Dow C S (2000) An electronic nose system for monitoring the quality of potable water. Sens Actuators B 69: 336
[63] Shykhon M E, Morgan D W, Dutta R, Hines E L, Gardner J W (2004) Clinical evaluation of the electronic nose in the diagnosis of ear, nose and throat infection: a preliminary study. J Laryngol Otol 118: 706
[64] Barko G, Hlavay J (1998) Application of principal component analysis for the characterisation of a piezoelectric sensors array. Anal Chim Acta 367: 135

[65] Guadarrama A, Fernandez J A, Iniguez M, Souto J, de Saja J A (2000) Array of conducting polymer sensors for the characterisation of wines. Anal Chim Acta 411: 193
[66] Guadarrama A, Fernandez J A, Iniguez M, Souto J, de Saja J A (2001) Discrimination of wine aroma using an array of conducting polymer sensors in conjunction with solid-phase micro-extraction (SPME) technique. Sens Actuators B 77: 401
[67] Jain A K, Murty M N, Flynn P J (1999) Data clustering: a review. ACM Computing Sur 31: 264
[68] Anderberg M R (1973) Cluster analysis for applications. Academic Press, Inc, New York
[69] Grate J W, Nelson D A, Skaggs R (2003) Sorptive behavior of monolayer-protected gold nanoparticle films: implications for chemical vapor sensing. Anal Chem 75: 1868
[70] Duda R O, Hart P E, Stork D G (2001) Pattern classification. Wiley-Interscience, New York
[71] Shaffer R E, Rose-Pehrsson S L, McGill R A (1999) A comparison study of chemical sensor array pattern recognition algorithms. Anal Chim Acta 384: 305
[72] Fisher R A (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7: 179
[73] Fisher R A (1938) The statistical utilization of multiple measurements. Ann Eugen 8: 376
[74] Anderson E (1935) The Irises of the Gaspe Peninsula. Bull Amer Iris Soc 59: 2
[75] Martin Y G, Oliveros M C C, Pavon J L P, Pinto C G, Cordero B M (2001) Electronic nose based on metal oxide semiconductor sensors and pattern recognition techniques: characterisation of vegetable oils. Anal Chim Acta 449: 69
[76] Di Natale C, Macagnano A, Martinelli E, Paolesse R, D'Arcangelo G, Roscioni C, Finazzi-Agro A, D'Amico A (2003) Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. Biosens Bioelectron 18: 1209
[77] Oliveros M C C, Pavon J L P, Pinto C G, Laespada M E F, Cordero B M, Forina M (2002) Electronic nose based on metal oxide semiconductor sensors as a fast alternative for the detection of adulteration of virgin olive oils. Anal Chim Acta 459: 219
[78] Goschnick J, Koronczi I, Frietsch M, Kiselev I (2005) Water pollution recognition with the electronic nose KAMINA. Sens Actuators B 106: 182
[79] Gonzalez-Martin I, Perez-Pavon J L, Gonzalez-Perez C, Hernandez-Mendez J, Alvarez-Garcia N (2000) Differentiation of products derived from Iberian breed swine by electronic olfactometry (electronic nose). Anal Chim Acta 424(2): 279
[80] Lavine B K (2002) Clustering and classification of analytical data. In: Meyers R A (eds) Encyclopedia of analytical chemistry. John Wiley & Sons Ltd, Chichester, p 1
[81] Schiffman S S, Wyrick D W, Gutierrez-Osuna R, Nagle H T (2000) Effectiveness of an electronic nose for monitoring bacterial and fungal growth. Olfaction and electronic noses, p 173
[82] Carmel L, Levy S, Lancet D, Harel D (2003) A feature extraction method for chemical sensors in electronic noses. Sens Actuators B 93: 67
[83] Bishop C M (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
[84] Jondarr G (1996) Backpropagation family album, Technical Report C/TR95-05
[85] Pardo M, Sberveglieri G (2004) Remarks on the use of multilayer perceptrons for the analysis of chemical sensor array data. IEEE Sensors J 4: 355
[86] Garcia-Gonzalez D L, Aparicio R (2003) Virgin olive oil quality classification combining neural network and MOS sensors. J Agric Food Chem 51: 3515
[87] Pardo M, Sberveglieri G, Gardini S, Dalcanale E (2000) A hierarchical classification scheme for an electronic nose. Sens Actuators B 69: 359
[88] Zou X B, Zhao J W, Wu S Y, Huang X Y (2003) Vinegar classification based on feature extraction and selection from tin oxide gas sensor array data. Sensors 3: 101
[89] Lowe D (1991) What have neural networks to offer statistical pattern processing? Proceedings of SPIE Conference on Adaptive Signal Processing 1565: 460
[90] Pedrycz W (1998) Conditional fuzzy clustering in the design of radial basis function neural networks. IEEE Trans Neural Networks 9: 601
[91] Byun H G, Persaud K C, Kim J D, Lee D D (1999) Improvement of RBF network's performance for application to odour sensing systems. Olfaction and electronic nose; The 6th international symposium Olfaction and electronic nose 99 and the 1st workshop of the Network of Excellence on Artificial Olfactory Sensing Nose, p 237
[92] Weymaere N, Martens J-P (1991) A fast and robust learning algorithm for feedforward neural networks. Neural Networks 4: 361
[93] Billings S A, Zheng G L (1995) Radial basis function network configuration using genetic algorithms. Neural Networks 8: 877
[94] Carse B, Fogarty T C (1996) Tackling the "curse of dimensionality" of radial basis function neural networks using a genetic algorithm. Lect Notes Comput Sci, p 710
[95] Looney C G (1997) Pattern recognition using neural networks. Oxford University Press, Oxford
[96] Distante C, Ancona N, Siciliano P (2003) Support vector machines for olfactory signals recognition. Sens Actuators B 88: 30
[97] Szczurek A, Maciejewska M (2004) Recognition of benzene, toluene and xylene using TGS array integrated with linear and non-linear classifier. Talanta 64: 609
[98] Ali Z, James D, O'Hare W T, Rowell F J, Scott S M (2003) Radial basis neural network for the classification of fresh edible oils using an electronic nose. J Therm Anal Calorim 71: 147
[99] James D, Scott S M, O'Hare W T, Ali Z, Rowell F J (2004) Classification of fresh edible oils using a coated piezoelectric sensor array-based electronic nose with soft computing approach for pattern recognition. T I Meas Control 26: 3
[100] Specht D F (1990) Probalistic neural networks. Neural Networks 3: 109
[101] McGill R A, Nguyen V K, Chung R, Shaffer R E, DiLella D, Stepnowski J L, Mlsna T E, Venezky D L, Dominguez D (2000) The "NRL-SAWRHINO": a nose for toxic gases. Sens Actuators B 65: 10
[102] Lozano J, Santos J P, Horrillo M C (2006) Classification of white wine aromas with an electronic nose. Talanta (in press)
[103] Santos J P, Fernandez M J, Fontecha J L, Lozano J, Aleixandre M, Garcia M, Gutierrez J, Horrillo M C (2006) SAW sensor array for wine discrimination. Sens Actuators B (in press)
[104] Kohonen T (1995) Self-organizing maps. Springer, Heidelberg
[105] Davide F A M, Di Natale C, D'Amico A (1994) Self-organizing multisensor systems for odour classification: internal categorization, adaptation and drift rejection. Sens Actuators B 18: 244
[106] Davide F A M, Di Natale C, Damico A (1995) Self-organizing sensory maps in odor classification mimicking. Biosens Bioelectron 10: 203

[107] Sinesio F, Di Natale C, Quaglia G B, Bucarelli F M, Moneta E, Macagnano A, Paolesse R, D'Amico A (2000) Use of electronic nose and trained sensory panel in the evaluation of tomato quality. J Sci Food Agr 80: 63
[108] Kohonen T (1989) Self-organization and associative memory, 3rd edn. Springer, New York
[109] Kusumoputro B, Budiarto H, Jatmiko W (2002) Fuzzy-neuro LVQ and its comparison with fuzzy algorithm LVQ in artificial odor discrimination system. ISA Trans 41: 395
[110] Seyama M, Nakamura M, Tate A (2004) Odor-sensor technology based on an array of quartz crystal resonators coated with plasma-deposited organic film. NTT Technical Review 2: 70
[111] Qian T, Xu R, Kwan C, Linnell B, Young R (2004) Toxic vapor classification and concentration estimation for space shuttle and international space station. Advances in neural networks – Isnn 2004, Pt 1. Springer, Berlin, p 543
[112] Zadeh L A (1965) Fuzzy sets. Information and Control 8: 338
[113] Zadeh L A (1968) Fuzzy algorithms. Information and Control 12: 94
[114] Bezdek J C (1981) Pattern recognition with fuzzy objective function algorithms. Plenum, New York
[115] Dutta R, Hines E, Gardner J, Boilot P (2002) Bacteria classification using cyranose 320 electronic nose. BioMedical Engineering On Line 1: 1
[116] Dutta R, Morgan D, Baker N, Gardner J W, Hines E L (2006) Identification of Staphylococcus aureus infections in hospital environment: electronic nose based approach. Sens Actuators B in press, corrected proof
[117] Dutta R, Kashwan K R, Bhuyan M, Hines E L, Gardner J W (2003) Electronic nose based tea quality standardization. Neural Networks 16: 847
[118] Marcelloni F (2001) Recognition of olfactory signals based on supervised fuzzy C-means and k-NN algorithms. Pattern Recognit Lett 22: 1007
[119] Dumitrescu D, Lazzerini B, Marcelloni F (2000) A fuzzy hierarchical classification system for olfactory signals. Pattern Anal Appl 3: 325
[120] Lazzerini B, Maggiore A, Marcelloni F (2001) FROS: a fuzzy logic-based recogniser of olfactory signals. Pattern Recogn 34: 2215
[121] Singh S, Hines E L, Gardner J W (1996) Fuzzy neural computing of coffee and tainted-water data from an electronic nose. Sens Actuators B 30: 185
[122] Scott S M, Ali Z, O'Hare W T (2003) Fuzzy logic and fuzzy classification techniques. In: Cartwright H M, Sztandera L M (eds) Soft computing approaches in chemistry. Springer, Heidelberg, p 95
[123] Grossberg S (1976) Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. Biological Cybernetics 23(3): 121
[124] Carpenter G A, Grossberg S, Reynolds J H (1995) A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems. IEEE Trans Neural Networks 6: 1330
[125] Carpenter G A, Grossberg S, Rosen D B (1991) Fuzzy art – fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4: 759
[126] Kasuba T (1993) Simplified fuzzy ARTMAP. AI Expert 8: 18
[127] Vakil-Baghmisheh M T, Pavesic N (2003) A fast simplified fuzzy ARTMAP network. Neural Processing Lett 17: 273
[128] Llobet E, Hines E L, Gardner J W, Bartlett P N, Mottram T T (1999) Fuzzy ARTMAP based electronic nose data analysis. Sens Actuators B 61: 183
[129] Vinaixa M, Marin S, Brezmes J, Llobet E, Vilanova X, Correig X, Ramos A, Sanchis V (2004) Early detection of fungal growth in bakery products by use of an electronic nose based on mass spectrometry. J Agric Food Chem 52: 6068
[130] Vinaixa M, Llobet E, Brezmes J, Vilanova X, Correig X (2005) A fuzzy ARTMAP- and PLS-based MS e-nose for the qualitative and quantitative assessment of rancidity in crisps. Sens Actuators B 106: 677
[131] Ali Z, Scott S M, James D, O'Hare W T, Rowell F J, Sarkodie-Guyan T, Theaker B J (2000) Classification of fresh edible oils with piezoelectric quartz crystal based electronic nose. In: Gardner J W, Persaud K C (eds) Electronic noses and olfaction. Institute of Physics Publishing, p 229
[132] Fritzke B (1994) Growing cell structures – a self-organizing network for unsupervised and supervised learning. Neural Networks 7: 1441
[133] Fritzke B (1995) A growing neural gas network learns topologies. Adv Neural Information Processing Syst 7: 625
[134] Marsland S, Shapiro J, Nehmzow U (2002) A self-organising network that grows when required. Neural Networks 15: 1041
[135] Cheng G J, Zell A (2001) Externally growing cell structures for data evaluation of chemical gas sensors. Neural Comput Appl 10: 89
[136] Llobet E, Hines E L, Gardner J W, Franco S (1999) Non-destructive banana ripeness determination using a neural network-based electronic nose. Meas Sci Technol 10: 538
[137] Vaid T P, Burl M C, Lewis N S (2001) Comparison of the performance of different discriminant algorithms in analyte discrimination tasks using an array of carbon black – polymer composite vapor detectors. Anal Chem 73: 321
[138] Bicego M, Tessari G, Tecchiolli G, Bettinelli M (2002) A comparative analysis of basic pattern recognition techniques for the development of small size electronic nose. Sens Actuators B 85: 137
[139] Dutta R, Hines E L, Gardner J W, Udrea D D, Boilot P (2003) Non-destructive egg freshness determination: an electronic nose based approach. Meas Sci Technol 14: 190
[140] Kuske M, Rubio R, Romain A C, Nicolas J, Marco S (2005) Fuzzy k-NN applied to moulds detection. Sens Actuators B 106: 52
[141] Saunders B W, Thiel D V, Mackay-Sim A (1995) Response kinetics of chemically-modified quartz piezoelectric crystals during odorant stimulation. Analyst 120: 1013
[142] Freeman N J, May I P, Weir D J (1994) Response kinetics of polymer-coated bulk acoustic wave devices on exposure to gases and vapours. J Chem Soc Faraday Transactions 90: 751
[143] Ratton L, Kunt T, McAvoy T, Fuja T, Cavicchi R, Semancik S (1997) A comparative study of signal processing techniques for clustering microsensor data (a first step towards an artificial nose). Sens Actuators B 41: 105
[144] Distante C, Leo M, Siciliano P, Persaud K C (2002) On the study of feature extraction methods for an electronic nose. Sens Actuators B 87: 274
[145] Llobet E, Brezmes J, Ionescu R, Vilanova X, Al-Khalifa S, Gardner J W, Barsan N, Correig X (2002) Wavelet transform and fuzzy ARTMAP-based pattern recognition for fast gas identification using a micro-hotplate gas sensor. Sens Actuators B 83: 238
[146] Leone A, Distante C, Ancona N, Persaud K C, Stella E, Siciliano P (2005) A powerful method for feature extraction and compression of electronic nose responses. Sens Actuators B 105: 378

[147] Skov T, Bro R (2005) A new approach for modelling sensor based data. Sens Actuators B 106: 719
[148] Grate J W (2000) Acoustic wave microsensor arrays for vapor sensing. Chem Rev 100: 2627
[149] Wilson D M, Dunman K, Roppel T, Kalim R (2000) Rank extraction in tin-oxide sensor arrays. Sens Actuators B 62: 199
[150] Carey W P, Beebe K R, Kowalski B R, Illman D L, Hirschfeld T (1986) Selection of adsorbates for chemical sensor arrays by pattern-recognition. Anal Chem 58: 149
[151] Lau K-T, Micklefield J, Slater J M (1998) The optimisation of sorption sensor arrays for use in ambient conditions. Sens Actuators B 50: 69
[152] Kermani B G, Schiffman S S, Nagle H T (1998) A novel method for reducing the dimensionality in a sensor array. IEEE Trans Instrum Meas 47: 728
[153] Park J, Groves W A, Zellers E T (1999) Vapor recognition with small arrays of polymer coated microsensors. A comprehensive analysis. Anal Chem 71: 3877
[154] Benedetti S, Mannino S, Sabatini A G, Marcazzan G L (2004) Electronic nose and neural network use for the classification of honey. Apidologie 35: 397
[155] Pardo M, Kwong L G, Sberveglieri G, Brubaker K, Schneider J F, Penrose W R, Stetter J R (2005) Data analysis for a hybrid sensor array. Sens Actuators B 106: 136
[156] Eklov T, Martensson P, Lundstrom I (1999) Selection of variables for interpreting multivariate gas sensor data. Anal Chim Acta 381: 221
[157] Polikar R, Shinar R, Udpa L, Porter M D (2001) Artificial intelligence methods for selection of an optimized sensor array for identification of volatile organic compounds*1. Sens Actuators B 80: 243
[158] Mitchell M (1996) An introduction to genetic algorithms. MIT Press, Cambridge M.A.
[159] Schiffman S S, Kermani B G, Nagle H T (1997) Analysis of medication off-odors using an electronic nose. Chem Senses 22: 119
[160] Kermani B G, Schiffman S S, Nagle H T (1999) Using neural networks and genetic algorithms to enhance performance in an electronic nose. IEEE Trans Biomed Eng 46: 429
[161] Pavlou A K, Magan N, Sharp D, Brown J, Barr H, Turner A P F (2000) An intelligent rapid odour recognition model in discrimination of Helicobacter pylori and other gastroesophageal isolates in vitro. Biosens Bioelectron 15: 333
[162] Corcoran P, Lowery P, Anglesea J (1998) Optimal configuration of a thermally cycled gas sensor array with neural network pattern recognition. Sens Actuators B 48: 448
[163] Corcoran P, Anglesea J, Elshaw M (1999) The application of genetic algorithms to sensor parameter selection for multisensor array configuration. Sens Actuators A 76: 57
[164] Boilot P, Hines E L, Gongora M A, Folland R S (2003) Electronic noses inter-comparison, data fusion and sensor selection in discrimination of standard fruit solutions. Sens Actuators B 88: 80
[165] Aleixandre M, Sayago I, Horrillo M C, Fernandez M J, Ares L, Garcia M, Santos J P, Gutierrez J (2004) Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas. Sens Actuators B 103: 122
[166] Llobet E, Brezmes J, Gualdron O, Vilanova X, Correig X (2004) Building parsimonious fuzzy ARTMAP models by variable selection with a cascaded genetic algorithm: application to multisensor systems for gas analysis. Sens Actuators B 99: 267
[167] Gardner J W, Boilot P, Hines E L (2005) Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach. Sens Actuators B 106: 114
[168] Kermani B G, Schiffman S S, Nagle H T (2005) Performance of the Levenberg-Marquardt neural network training method in electronic nose applications. Sens Actuators B 110: 13
[169] Davide F A M, Di Natale C, D'Amico A, Hierlemann A, Mitrovics J, Schweizer M, Weimar U, Gopel W, Marco S, Pardo A (1995) Dynamic calibration of QMB polymer-coated sensors by Wiener kernel estimation. Sens Actuators B 26: 275
[170] Fryder M, Holmberg M, Winquist F, Lundstrom I (1995) A calibration technique for an electronic nose. Conference Solid-State Sens Actuator 683
[171] Sisk B C, Lewis N S (2005) Comparison of analytical methods and calibration methods for correction of detector response drift in arrays of carbon black-polymer composite vapor detectors. Sens Actuators B 104: 249
[172] Holmberg M, Winquist F, Lundstrom I, Davide F, DiNatale C, D'Amico A (1996) Drift counteraction for an electronic nose. Sens Actuators B 36: 528
[173] Holmberg M, Davide F A M, Di Natale C, D'Amico A, Winquist F, Lundstrom I (1997) Drift counteraction in odour recognition applications: Lifelong calibration method. Sens Actuators B 42: 185
[174] Gardner J W, Hines E L, Pang C (1996) Detection of vapours and odours from a multisensor array using pattern recognition: self-organising adaptive resonance techniques. Meas Control-UK 29: 172
[175] NOSE II (2005) Network on artificial Olfactory Sensing, http://www.nose-network.org/