Machine Learning: a crucial tool for sensor design

Weixiang Zhao, *University of California, Davis, CA, USA*
Abhinav Bhushan, *of California, Davis, CA, USA*
Anthony D. Santamaria, *University of California, Davis, CA, USA*
Melinda G. Simon, *University of California, Davis, CA, USA*
Cristina E. Davis, *University of California, Davis, CA, USA*

ABSTRACT. Sensors have been widely used for disease diagnosis, environmental quality monitoring, food quality control, industrial process analysis and control, and other related fields. As a key tool for sensor data analysis, machine learning is becoming a core part of novel sensor design. Dividing a complete machine learning process into three steps: data pre-treatment, feature extraction and dimension reduction, and system modelling, this paper provides a review of the methods that are widely used for each step. For each method, the principles and the key issues that affect modelling results are discussed. After reviewing the potential problems in machine learning processes, this paper gives a summary of current algorithms in this field and provides some feasible directions for future studies.

KEYWORDS: *Feature extraction, Machine learning, Multivariate analysis, Neural networks, Optimization, Sensors*

Introduction

Sensors have been applied to a variety of fields such as environmental quality monitoring (Roncaglia et al., 2004), non-invasive disease diagnosis (Chen et al., 2005), food quality control (Marin et al., 2007), and industrial process analysis (Yan, Zhao, 2009), because of the following advantages: (1) ability to function in a harsh environment, (2) ability to operate continuously and automatically, and (3) high accuracy and sensitivity. In general, developing a sensor depends on two major components: analytical instruments and data analysis techniques. Novel analytical instruments allow producing a great amount of information (data) and also permit the exploration of new fields.

However, these generated sensor data may contain irrelevant information and moreover the principles of the new fields could be very complex and even totally unknown, so reliable sensor systems

are becoming increasingly reliant on sophisticated data processing techniques.
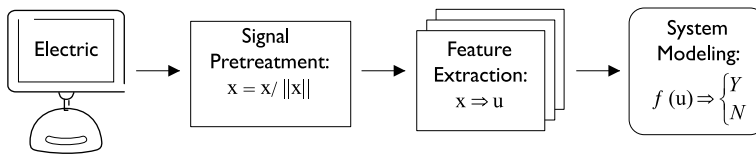
Figure 1.
Flowchart of a machine learning process for electric sensor systems.

 As a powerful tool for advanced data processing, machine learning has become a core technique for novel sensor development, aiming to discover the hidden rules that control complex systems. As shown in Figure 1, a complete machine learning process is composed of three steps: data pre-processing, feature extraction and dimension reduction, and system modelling. Data pre-processing comprises noise filtering, data normalization, signal alignment, and other related data treatments. Sensor signals are usually composed of a large number of variables, so in the second step feature extracting methods are used to transfer sensor signals from their original high dimensional space to a low dimensional feature space or to select "representative" (pertinent) variables to characterize an entire system. Given a good feature expression, the last step of machine learning is to establish system models either for classification problems such as pollutant detection and disease diagnosis or quantitative estimation problems like chemical concentration prediction.

Any of these three steps can play a key role in controlling machine learning effects. There have been some good review papers and books about pattern recognition techniques for either general industrial applications or specific fields like food science (Bishop, 2006; Berrueta et al., 2007). This paper will provide a review of the algorithms that are currently used in each step, compare their individual properties, and discuss future perspectives of machine learning methods for sensor development, covering the applications to both classification and quantitative estimation.

Data pre-processing

Data pre-treatment is important and also the first step for establishing a machine learning model, as raw sensor signals usually (even unavoidably) have some problems that could be harmful for modelling effects.

Noise removal

Noise removal is a basic procedure for signal enhancement. Many sensor signals are composed of time series data, so a variety of digital signal processing and time series analysis techniques have been applied for signal enhancement. Different from Fourier analysis which only focuses on frequency domain, wavelet analysis examines data in both time and frequency domains. It has become a powerful noise filtering method, decomposing original signal into low and high frequency domains (Addison, 2002). The high frequency domain contains more noise information from the original data, so modifying the wavelet coefficients in the high frequency domain by setting up a threshold is a simple but effective method for sensor signal enhancement (Cappadona et al., 2008). Auto-regressive (AR) analysis is another noise filtering technique, generating AR coefficients to represent and reconstruct original signal. It has been successfully used for chromatographic signal enhancement (Krebs et al., 2005; Ubeyli, Guler, 2004).

Baseline removal

Baseline removal is another important pre-treatment for signal enhancement. A conventional baseline removal process for spectral or time series data consists of three major steps (Pearson, 1977): (1) to determine baseline key points in spectrum, (2) to build a baseline model for the whole spectrum using the detected baseline points, and (3) to correct the signal by subtracting the baseline from original signal. Recently, some new algorithms have been developed, such as an adaptive learning algorithm for electrocardiogram (ECG) baseline removal (Esposito, D'Andria, 2003) and a selective filter for ECG baseline correction (Shusterman et al., 2000).

Signal alignment

Signal shifting is a potential problem for data series. Chromatography/ spectral instrument based sensors are a typical place where this

problem often occurs. It is no doubt that shifted signals can result in a failure of sensor analysis. In general, for chromatography data, time alignment begins with locating the peaks that correspond to the same chemicals in a selected template chromatography profile. Then, after getting these peaks aligned, a linear or nonlinear interpolation process is applied for chromatography profile registration. Spline functions are an efficient nonlinear interpolation method (Krebs, Tingley et al., 2006). Figure 2 shows a significant improvement on the separation of two groups of gas chromatography mass spectrometry (GC/MS) data before and after time alignment process (Krebs, Tingley et al., 2006), by employing a clustering tool called principal component analysis. The details of this tool will be discussed in the next section. As an extension of one-dimensional alignment, two dimensional alignment can be applied to the three dimensional spectral data, such as gas chromatography differential mobility spectrometry (GC/DMS) data (Krebs, Kang et al., 2006).
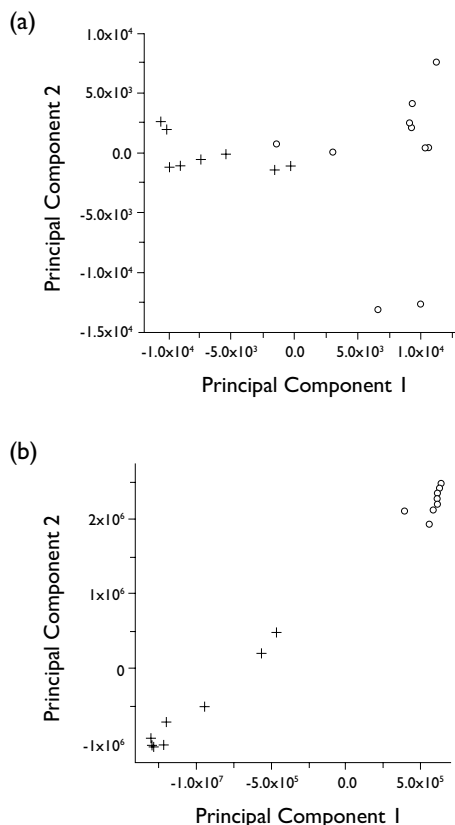

Figure 2.
Principal component analysis of total ion chromatograms resulting from GC–MS analysis of headspace above plasma samples from two donors (Donor A: +, Donor B: o).
(a) before and (b) after alignment.

Outlier detection

To have a reliable and automated sensor system, data pre-treatment also needs to detect possible outliers that could influence modelling effects. Unlike noise, an outlier is an observation that disobeys the distribution of the rest of the data (Crowe, 1989). Outliers can also be called "questionable data", "strange points", or "doubtable observations", and are often caused by confounding system mechanisms or instrument error (Mah, Tamhane, 1982). Outliers may cause damage to a modelling system to which they do not belong, so a broad range of techniques have been applied for outlier detection. Statistical analysis methods are widely used to detect outliers in linear systems, estimating the standardized residual of each observation in the regression model (Mah, Tamhane, 1982; Prescott, 1975). Self-organizing map (SOM) was introduced to detect outliers in nonlinear systems, based on the distance between data points (Munoz, Muruzabal, 1998). However, in some cases, distance is not a proper criterion for outlier verification, because some normal data may deviate more from the majority of the data than outliers, especially for highly nonlinear systems. A method integrating a linear outlier detection strategy with radial basis function neural networks (RBFNN) was proposed to detect outliers in complex nonlinear systems (Zhao, Chen, Hu, 2004). Outliers are not noise, so simply excluding outliers from analysis may be not beneficial to the discovery of the hidden mechanisms and rules.

Data normalization

Normalization usually is the final step for data pre-treatment. In terms of data characteristics, this step can be divided into local normalization and global normalization. A typical local normalization process is unit scaling, which is widely used for spectral data. The similarity of two spectral samples is represented by the dot product of their unit spectral vectors. Global normalization aims to make sensor variables comparable, preventing some variables with lower magnitudes from being overwhelmed by others with higher magnitudes. However, in some cases, global optimization can amplify the noise or the information of irrelevant sensor variables, which might be harmful for modelling.

Feature extraction and dimension reduction

The development of novel instruments has produced sensors that generate data with extremely high dimensions, which provides an opportunity to carry out an extensive system investigation. However, analyzing these copious amounts of data depends on sophisticated techniques to extract pertinent features from original data. Generally, feature extraction and dimension reduction aim to create a map from a high dimensional space to a low dimensional space $f : R^m \rightarrow R^n (n << m)$. This step not only unloads heavy computational burdens from the subsequent system modelling step but also excludes irrelevant and noise signals from analysis.

Traditional multivariate analysis methods

Multivariate statistical analysis is a very efficient strategy for feature extraction. Principal component analysis (PCA) is the most widely used method for this purpose. PCA is a linear transformation that transforms original data to a new coordinate system where the greatest variance of the data lies on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so forth. In other words, PCA is an eigenvector-based multivariate analysis method, providing a map from an original data space to a new space with much lower dimensions (Wold et al., 1987). With the abilities to remove co-linearity in variables, condense information of original data, and enhance signal quality, PCA has been extensively applied for sensor signal analysis such as fungal growth detection in bakery products (Vinaixa et al, 2004), olive oil discrimination (Brezmes et al., 2005), and environmental quality monitoring (Scorsone et al., 2006). As extension of PCA, independent component analysis (ICA) is another effective feature extraction method, extracting statistically independent components from original data space (Comon, 1994; Krier et al., 2008). Different from PCA that solely de-correlates the data, ICA has obtained great applications to sensor design (Yadava, Chaudhary, 2006; Di Natale et al., 2002).

Linear discrimination analysis (LDA) and partial least square regression (PLSR) are also widely used as statistical feature extraction methods for sensor development (Buratti et al., 2004; Vestergaard et al., 2006). Different from PCA, the extractions based on LDA and PLSR are embedded in a classification or regression process that requires both

independent data X and dependent data Y. The principles of these two methods will be discussed in the following "System modelling" section.

Positive matrix factorization is another important feature extraction method, especially in the cases where the generated "loading" and "score" matrixes should only have positive values to ensure their physical meanings. A positive matrix factorization (PMF) method based on least square regression has been extensively applied to air quality analysis and control (Paatero, 1997; Zhao, Hopke, 2004; 2006), with potential applications to a variety of environmental sensor design.

Digital signal processing based methods

In many cases, sensor output signal is a time series, such as a chromatogram, or an image, such as mass spectrometry - mass spectrometry (MS/MS) data, so digital signal processing and time series analysis have become powerful feature extraction methods for sensor signals.

Wavelet analysis

In contrast to the traditional fast Fourier transformation, wavelet analysis is able to examine the signal simultaneously in both time and frequency domains, so it is an excellent tool to analyze non¬periodic, noisy, and intermittent signals and has spawned a number of wavelet-based methods for signal analysis and interrogation (Addison, 2002). Basically, wavelet transformation aims to represent an arbitrary function by superposing a group of wavelets which are generated from a mother wavelet Ψ through dilations and translations. A wavelet function generated at scale a and location b can be described as,

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \qquad (1)$$

Proposed by Mallet for digital signal processing, wavelet analysis was substantially developed for various application fields after Daubechies constructed a set of wavelet orthonormal basis functions (Mallet, 1989; Daubechies, 1988). In a recent study of sensor selection for machine olfaction design (Phaisangittisagul, Nagle, 2008), discrete wavelet transformation not only significantly

reduced the number of sensor variables but also yielded an almost 100% accuracy for the classification of two types of odour (coffee and soda). Wavelet transformation coupled with artificial neural networks was also successfully used for electronic tongue design (Moreno-Baron et al., 2006) and environmental variable monitoring (Ciarlini, Maniscalco, 2008). To ensure the feature extraction effect of wavelet analysis, a couple of things should be taken into account including wavelet type and decomposition level. The coefficients of an overly-deep decomposition level might not have enough signal information, while those of a low decomposition level may still contain much noise. Determining how many levels of wavelet coefficients are needed and whether the coefficients in both low and high frequency domains are useful is case-dependent.

Auto-regressive modelling
Auto-regressive (AR) analysis is an efficient feature extraction tool for time series data, widely used for speech analysis (Markovic et al., 2002; Smidl, Quinn, 2005). As an all pole model (filter), a p-order AR model can be expressed as:

$$x_n = a_o + \sum_{i=1}^{p} a_i x_{n-i} + \varepsilon_n \qquad (2)$$

where x is time series data, ai (i=1, ...,p) are AR coefficients, a_0 is a constant, and ε_n is an error estimate. Briefly, using AR coefficients, we can represent the nth value xn with its previous p values: x_{n-1}, x_{n-2}, ..., x_{n-p}. The goal of an AR model is to estimate the AR coefficients that can fit the original data series as closely as possible through an optimization process.
AR coefficients systematically characterize a changing trend in the data series, so in addition to the common effects of signal treatment, such as noise removal and dimension reduction, the AR model shows its significant advantage in dealing with mild signal shifts in the time series (such as chromatograph data). AR modelling can permit us to directly work on time shifted chromatograph data without using time alignment as a pre-processing step. This advantage has recently been proven in a GC/MS based sensor design for bacterial identification (Zhao, Morgan, Davis, 2008).

Feature subset selection

The output of sensor array or even a single sensor can be composed of a high dimensional vector, with each element representing a physical variable or a chemical compound. However, it is very likely that only some of these variables contain pertinent information, so feature subset selection not only reduces signal dimension but also excludes irrelevant variables from system modelling. More importantly, the selected feature subset can provide the most direct and pertinent information for system analysis. For example, detecting biomarkers from the signals of a biosensor designed for disease diagnosis can provide the kernel information for pathology research.

In general, selecting a feature subset is an optimization problem which aims to find a subset that can yield an optimal solution such as the highest classification accuracy. Because the objective function could be represented by non-traditional models like neural network models, stochastic optimization methods such as genetic algorithms (Goldberg, 1989), differential evolution (Storn, Price, 1995), and simulated annealing (Kirkpatrick et al., 1983) are popularly used in this aspect.

Genetic algorithm (GA)

GA is a powerful optimization method that mimics natural evolution principles, consisting of three major operators: selection, crossover, and mutation (Michalewicz et al. 1992; Hibbert, 1993). By employing an objective function constructed by selected variables, GA attempts to provide an optimal feature subset. In this optimization problem, a binary vector indicating which variables are selected into the feature subset is used as chromosome for the evolution process in GA and the objective function represented by classification or predication accuracy is used as a fitness function to evaluate chromosome quality.
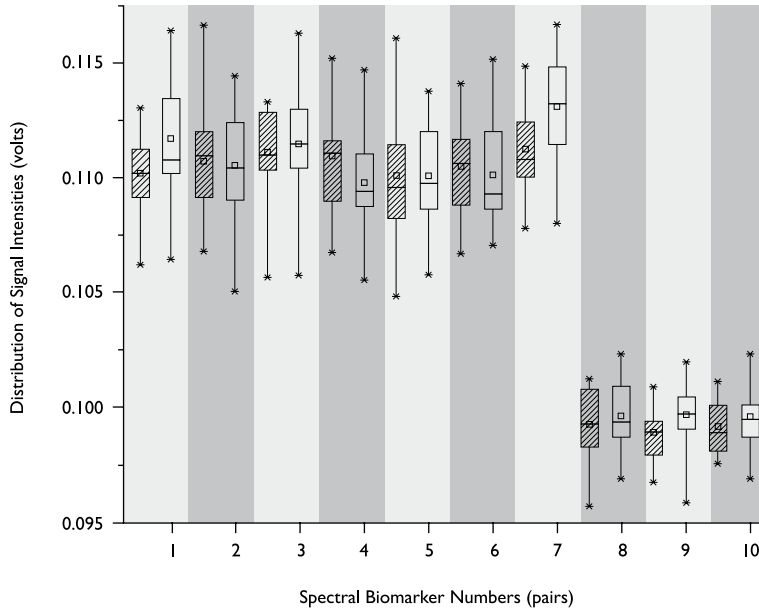
Figure 3 shows 10 biomarkers that were detected by GA to differentiate protein and virus from the output signals of a gas chromatography differential mobility spectrometry (GC/DMS) based sensor (Ayer et al., 2008). Using a PCA and neural network integrated classification model, these 10 biomarkers yielded 94% classification accuracy. The GA- driven feature subset selection approach was also applied for DMS based bacterial identification and Bacillus spore detection (Krebs, Mansfield et al., 2006; Shnayderman et al., 2005) and MS based electric nose design (Llobet et al., 2007). Although GA can theoretically provide a global optimal solution, for a high dimensional optimization problem, GA is very likely to fall into local optima, which usually indicates a premature solution. It is therefore a key issue to define proper cross-over and mutation rates, although it can be difficult in some cases.

Differential evolution (DE)
DE is a novel parallel optimal searching method proposed in the mid-1990s (Storn, Price, 1995). The main concept of DE is a scheme to generate population vectors. Basically, DE employs three population vectors to generate a new population vector. The weighted difference between two selected population vectors

is added to the third vector, which is then crossed-over with the target vector g to generate a new vector v. If v has a better quality than g, it will replace g for the next generation, otherwise, v will be discarded and g will survive (Li, Heinemann, 2007). This "family" generation process is different from that of GA, where two randomly selected members are used to generate a member for the next generation.

Using a probabilistic neural network (PNN) model as the cost function, DE was applied for the wavelength selection to develop a surface acoustic wave sensor for food quality monitoring (Li, Heinemann, 2007). In this study, a comparison between the results of DE and GA shows that although both methods significantly reduced data dimensions by 50%, DE achieved a better average performance than GA, reaching the minimal misclassification rate of 0.0175.

Simulated annealing (SA)

Simulated annealing (SA) is another global searching approach motivated by statistical thermodynamics (Kirkpatrick et al., 1983). During an optimization process, SA algorithms replace the current solution with a random neighbour solution based on a probability p which is a function of current "annealing" temperature (T). This replacement can help prevent searching processes from falling into local optima. SA has been applied to detect "representative" mass spectrometry fragments for an E-nose based food classification (Llobet et al., 2007). The key for SA is to establish a proper "annealing" temperature (T) decrease function. In some cases, it could be challenging to design a proper decreasing process of T and also SA could require prohibitively long periods of time to follow this process.

System modelling

With the data pre-processing and feature extraction steps done, a machine learning process is moving to the final step, a real learning process called system modelling. In terms of learning strategies, the learning processes can be divided into two categories, supervised leaning and unsupervised learning.

Supervised learning methods

Assuming the sensor output signals of the analyzed samples are $\mathbf{X}$ and the corresponding information (e.g., class memberships or chemical concentrations) are $\mathbf{Y}$, supervised learning process aims to establish a function $f:\mathbf{Y}=f(\mathbf{X})$ to describe the relationship between $\mathbf{X}$ and $\mathbf{Y}$ and make a prediction for a new sample. Most supervised learning methods can be used for both classification problems like vapour identification and quantitative estimation problems like vapour concentration prediction, while some can be used only for classification problems.

Principal component regression (PCR)
As a simple machine learning method, least square (LS) regression is the most popular method to create a map from independent variables $\mathbf{X}$ to dependent variables $\mathbf{Y}$. The major problems for ordinary LS methods are that (1) the sample number must be equal to or larger than the number of model parameters (i.e., the number of independent variables) and (2) possible co-linearity in the independent variables could result in the ill-condition of a regression matrix and eventually lead to an unreliable solution. One feasible way to solve these problems is to create a regression model based on principal components, called PCR (Wold et al., 1987). Thus, the regressors are decorrelated and the number of regressors is also significantly reduced. However, the extraction of principal components is solely reliant on the independent variables (sensor signals), so the direction of maximal variance in PCA is not related with the dependent variables, which usually results in the modelling effect of PCR being a little poorer than the effect of a model to be discussed next.

Partial least square regression (PLSR)
PLSR is another modified LSR, extracting uncorrelated latent variables from the original data. The difference between PLSR and PCR is the extraction of latent variables in PLSR not only employs independent variables $\mathbf{X}$ but also take into account dependent variables $\mathbf{Y}$.
A general linear model is $\mathbf{Y}=\mathbf{XB}+\mathbf{E}$, where $\mathbf{B}$ is the regression coefficients and $\mathbf{E}$ is the residual matrix. PLSR generates a transforming matrix $\mathbf{W}$ to transform $\mathbf{X}$ to $\mathbf{T}$ ($\mathbf{T}=\mathbf{XW}$, called score matrix or PLS components). Let $\mathbf{\Lambda}=\mathbf{WB}$, then the PLS regression

model turns to be $\mathbf{Y}=\mathbf{T\Lambda}+\mathbf{E}$. Thus, the regression of $\mathbf{X}$ against $\mathbf{Y}$ becomes the regression of PLS components $\mathbf{T}$ against $\mathbf{Y}$. One standard PLS algorithm is nonlinear iterative partial least squares (NIPALS). The details of this algorithm are discussed in the literature (Hoskuldsson, 1988).

Both PCR and PLSR were designed for quantitative estimation problems and classification problems as well. As an effective linear function approximation tool, PLSR has been widely applied to E-nose design (Vestergaard et al., 2006; Trihaas et al., 2005; Aishima, 2004).

Linear discrimination analysis (LDA)

LDA is a linear classifier that expects to find the best separating line or plane between two groups of samples. Figure 4 shows a schematic LDA based separation of two groups. Assuming a sensor output vector is $\mathbf{x}$, a general equation for LDA can be expressed as:

$$f(x) = w^T x + w_O \qquad (3)$$

where $\mathbf{w}$ is a weight vector and w_0 is a constant (threshold value). The decision rule for a two-class problem is based on the positivity or negativity of the function value of $f(\mathbf{x})$.

Figure 4.
A schematic illustration of a two-class linear separation.
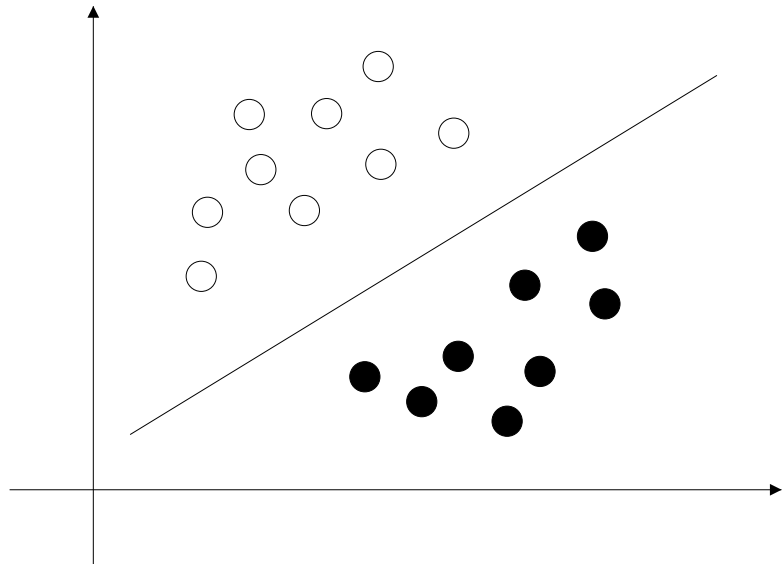
The key for designing a LDA classifier is to obtain the best separation line or plane (i.e., the best $\mathbf{w}$ and w_0) from the training samples. Fisher discrimination (Fischer, 1936) is the most canonical strategy for finding the best parameters $\mathbf{w}$ and w_0. Given the between - and within-covariance matrices (SB and SW) of two sample groups, Fisher discrimination is to maximize the ratio of the between-to-within variability through the following optimization problem.

$$\max J(w) = \frac{w^T S_B w}{w^T S_W w} \qquad (4)$$

The above classification strategy can be extended to a multiple-class separation problem. Stepwise linear discrimination functions are a feasible tool for such an extension. LDA has also been extensively used in E-nose applications such as wine classification (Buratti et al., 2004) and green tea grade identification (Yu et al., 2008).

Bayes classifier

Bayes classifier is a maximal probability rule based classification method, aiming to minimize the "expected risk" caused by misclassification (Dutta, Dutta, 2006). Assuming there are N classes (1,2,.., N), the posterior probability of class h for sample x is:

$$P(h/x) = \frac{\lambda_h P(x/h)}{\sum_{i=1}^{N} \lambda_j P(x/j)} \qquad (5)$$

where, λ_j is the priori probability of class j and $P(\mathbf{x}/j)$ is the conditional probability of $\mathbf{x}$ given class j.
The Bayes decision rule is:

$$w^{opt} = \sum_i \alpha_i^{opt} y_i x_i \cdot w^{opt} \qquad (6)$$

The priori probability λ_h can be estimated as the ratio of the number of the samples in Class h to the number of all the samples. In most cases, the conditional probability $P(\mathbf{x}/h)$ can be replaced by a continuous probability density function $f(\mathbf{x}/h)$ and usually can be estimated by two methods, namely the parametric approach and non-parametric approach (Dutta, Dutta, 2006). A Bayes classifier based on parametric approach can also be called quadratic

discrimination (QD) process while non-parametric approaches generally apply kernel functions to estimate $f(\mathbf{x} / h)$.

Bayes classifier was successfully applied to classify the bacterial infections in a hospital environment (Dutta, Dutta, 2006). In this study, using an adaptive kernel function based probability density function, the Bayes classifier reached an excellent bacterial classification accuracy of 99.8%.

k-nearest neighbours (k-NN)

k-NN is a relatively straightforward pattern recognition method for sensor design (Dodd et al., 2004; Kuske et al., 2005). Assuming there are N samples belonging to C classes, in order to classify a new sample, k-NN needs to find k nearest neighbours of this new sample from the N samples based on distance or similarity. Let the k nearest samples be composed of k_1 from class 1, k_2 from class 2, …, and k_C from class C. Thus, the decision rule is:

$$x = \arg\max_{j} : k_j \quad (j = 1, 2, ..., C) \qquad (7)$$

The new sample belongs to the class that has the most neighbours of this sample. The key parameter for this method is k. A larger value of k could reduce the influence of outlier or noise but could generate clouded class boundaries. Usually small k values (3 or 5) are preferred (Berrueta et al., 2007).

The advantages of k-NN are (1) mathematical simplicity and (2) no need for statistical assumptions for sample distribution. However, k-NN may not work well when the sample numbers of each class are not comparable. In such cases, using weighted distance for clustering is a feasible choice. Meanwhile, considering the determination of k could be empirical and arbitrary, an adaptive k-NN was recently proposed to provide a robust categorization result (Roncaglia et al., 2004), where the categorization of a new sample is not solely based on the number of its neighbours in each class. This adaptive k-NN approach presented a good application in detecting air pollutants with a sensor array (Roncaglia et al., 2004).

Back-propagation neural network (BPNN)

From this subsection on, the discussion of supervised learning methods will focus on artificial neural network (ANN) models. Some

modifications of linear learning methods such as stepwise functions and polynominal regression can be applied for non-linear systems, but they do not work for highly nonlinear modelling problems. Presenting a good mimic of human brain cognition process, ANN has been extensively used for machine learning tasks. A variety of ANN models have been developed for sensor design, but this paper will focus on three most widely used ANN models, since many of others can be considered derivatives of these versions.

BPNN is one of the most widely used neural network models, with extensive applications in function approximation (classification is also a type of function approximation problem). Typically BPNN is composed of three layers: input, hidden, and output. A sigmoid function is a standard nonlinear activation function for each hidden and output neuron. Given the error between the stipulated and predicted results for the kth sample, Ek, the learning process (i.e., the adjusting process for connecting weights) based on the gradient descent algorithm can be represented as:

$$\Delta v = \alpha \left[\frac{\partial E_k}{\partial v} \right] \ and \ \Delta w = -\beta \left[\frac{\partial E_k}{\partial v} \right] \qquad (8)$$

where, α and β are the learning rates, $\mathbf{v}$ and $\mathbf{w}$ are the connecting weight vectors between hidden and input layers and between output and hidden layers, respectively. This learning process is repeated for all the training samples, which is called one iteration. This iteration process continues until a convergence criterion is reached.

BPNN has been used for developing various E-nose systems including bacterial infected illness diagnosis and warfare agent stimulants classification (Monge et al. 2004; Tchoupo, Guiseppi-Elie, 2005; Gardner et al., 2000; Alizadeh, Zeynali, 2008). A three-layer BPNN with sufficient hidden neurons has proven able to approximate any classification boundaries (Cybenko, 1989), but a possible over¬fitting problem will lead to a failure in the prediction of new samples. A constructive approach for building BPNN could be either starting with a small structure followed by a gradual increase of hidden neurons or using pruning strategies to remove extra hidden neurons from a big initial model structure.

Possible overlong learning time and local optima are two major problems for BPNN. Adaptive learning strategies can help BPNN

reach a convergence and jump out of local optima, but designing such a self-adjusting process is empirical and difficult.

Radial basis function neural network (RBFNN)
RBFNN is another widely used neural network model. The basic concept of RBFNN is a radial function-based interpolation process, seeking a function $f(\mathbf{x}_i) = \mathbf{y}_i$ to create a map from sample set $\{\mathbf{x}_i\}$ to their corresponding outputs $\{\mathbf{y}_i\}$. The typical structure of RBFN is also a three-layer forward structure. The hidden layer performs a nonlinear transformation to transform the input space into a high dimensional transitional space through radial basis functions. Gaussian kernel functions are a typical choice of radial basis functions.

$$\varphi(x, c_j) = \exp(- \| x - c_j \|^2 / \sigma_j^2) \qquad (9)$$

where $\mathbf{x}$ is a learning sample vector, $\mathbf{c}_j$ is the radial basis vector of the jth hidden node, σ_j is the Gaussian width of the jth hidden node. With the weights connecting the output node and the hidden nodes, the output layer produces a linear summation of all the hidden layer outputs. RBFNN can avoid long training time, performing a one-step learning strategy.
RBFNN was successfully applied for bacteria classification (Dutta et al., 2006), yielding a 98% classification accuracy for six bacteria species. RBFNN was also used to develop sensors for odour classification (Kim et al., 2007) and fragrance discrimination (Branca et al., 2003).
The major hindrance of building up an RBFNN model is to estimate the radial basis vectors and their corresponding Gaussian widths. A conventional strategy to determine the radial basis vector is to use K-means algorithm to group training samples into a number of clusters and then use the cluster centre vectors as the radial basis vectors (Hush, Horne, 1993). However the possible local optima of K-means algorithm could make the cluster centres unable to correctly reflect the data space distribution. In the early 1990's, orthogonal least squares (OLS) was introduced to construct RBFNN, using a Gram-Schmidt scheme to select regressors (i.e., radial basis vectors) from the training data (Chen et al., 1991). Genetic algorithms were also applied to determine which training samples could be used as radial basis vectors by minimizing the training error

(Whitehead, Choate, 1996), but the searching process may take a very long time and can be stuck in local optima. Integrating RBFNN with multivariate analysis methods is another effective strategy for radial basis vector construction (Walczak, Massart, 1996). In the RBF-PLSR approach, all the training samples are used as radial basis vectors and then PLSR is applied for the linear regression between the hidden and output layers. This method statistically solves the construction of radial basis vectors and has been used for mass spectrometry calibration and GC/MS based bacterial classification (Zhao et al., 2005; Zhao, Morgan, Davis, 2008).

Support vector machine (SVM)
The principle for SVM was proposed in the 1960s, but extensive studies and applications of SVM began in the 1990s. Principally SVM is a statistical method, but in many cases it can be categorized in the neural network field, as it has the same structure as RBFNN. In contrast to BPNN which requires defining a number of hidden neurons before training, the structure of SVM is adaptively determined during a learning process. Figure 5 shows a schematic linear classification based on support vectors. The black circles and triangles are the support vectors of two classes, respectively. Basically, SVM is to find a discrimination function as shown in equation (10), which is the same as LDA, but the calculation of the weight vector is only based on the support vectors.

$$f(\mathrm{x}_i) = w^T x_i + b \qquad (10)$$

where $\mathbf{x}_i$ is a sample vector, $\mathbf{w}$ is the weight vector and b is a threshold value.
Assuming support vectors have been determined, the optimal separation plane is the one that can maximize the margin between two classes, which is proved to be $2 \, / \, ||\mathbf{w}||$ (Kecman, 2001). Introducing Lagrange multipliers to this quadratic optimization problem can yield,

$$\min : L(w,b,\alpha) = \frac{1}{2} w^T w - \sum_i \alpha_i \{ y_i [w x_i + b] - 1 \} \quad (11)$$

Figure 5.
Support vectors based classification system.

Margin: $\dfrac{2}{\|\mathrm{w}\|}$
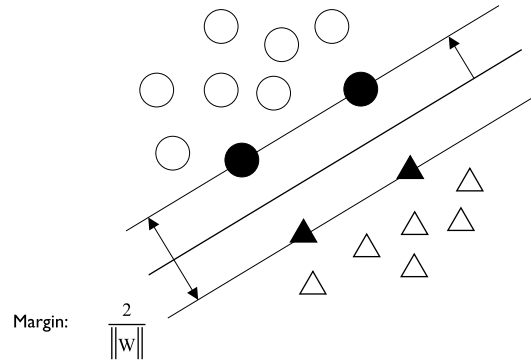
where α_i are Lagrange multipliers. The optimal solution is

$$w^{opt} = \sum_i \alpha_i^{opt} y_i x_i \cdot w^{opt}$$

can be easily calculated by any support vector using $\mathbf{w}^{opt}$. Support vectors are the samples whose $\alpha_i \neq 0$. The separation process can be extended to multiple-class problems. Using Gaussian kernel functions to transform the original input space to a linear transitional space, the linear separation process of SVM can be easily extended to nonlinear systems. Gaussian kernel functions are also a key issue for RBFNN, so in some cases SVM for nonlinear systems can also be considered as a specific modification of RBFNN where only the support vectors (samples) are used for constructing a classifier. SVM has obtained successful applications in many fields such as E-nose based vapours detection for environmental monitoring (Qian et al., 2006) and alcohol identification (Acevedo et al., 2007). Using Gaussian kernel function for signal space transformation, SVM also presents some inherent problems including the determination of a proper Gaussian width for each kernel function. In addition to the above three ANN models, probability neural network (PNN) is also popularly used for sensor signal analysis. Basically, PNN has the same principle as Bayes classifier. PNN also used kernel functions for signal space transformation, so PNN can be considered another specific modification of RBFNN or a neural network implementation of Bayes classifier. Kernel functions have been also integrated with PLS and PCA to extend the application fields of these two classical machine learning tools from linear problems to nonlinear

problems. RBF-PLS discussed above is a typical kernel PLS model.
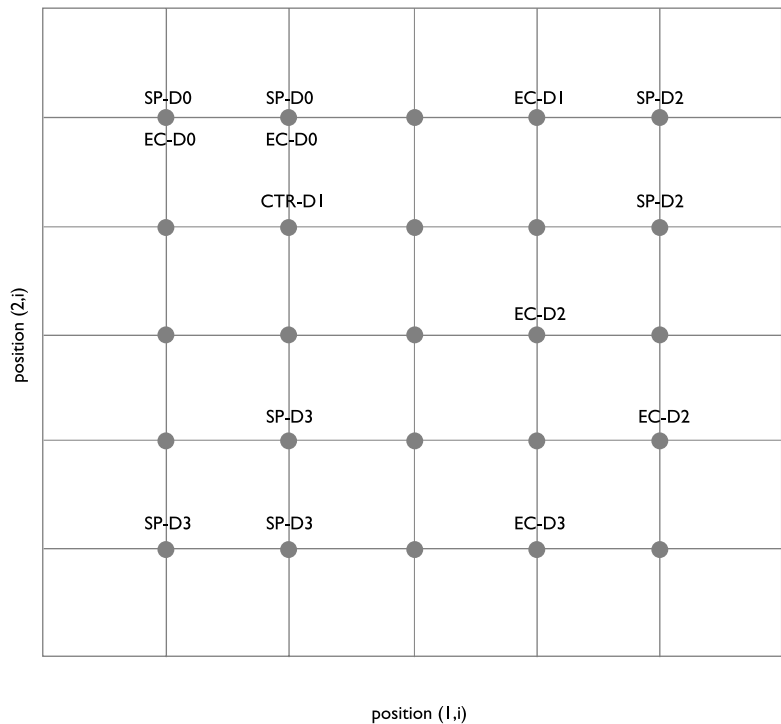
Unsupervised learning methods

Unsupervised learning is used to examine how the data are organized. In contrast to supervised learning processes, unsupervised learning processes are only given unlabeled samples. In other words, learners only have independent information $\mathbf{X}$ but no corresponding information $\mathbf{Y}$. Cluster analysis is a major topic for unsupervised learning methods. For sensor signal analysis, unsupervised learning methods can not only display the distribution and the grouping of the data but also provide concentrated information for supervised learning processes.

PCA can be considered as the simplest unsupervised learning method, providing a visual data grouping result. Another simple but effective unsupervised learning method is K-means algorithm. The basic concept of this algorithm is to minimize the sum of the distances between all the samples to their cluster centres (i.e., total intra-cluster variance) through an iteration process. K-means algorithm has been widely used as a tool for system resolution and feature extraction. For example, K-means is a practical method to determine radial basis vectors for RBFNN and centre vectors for PNN. The major drawbacks of K-means are that (1) the global convergence of this algorithm largely depends on the initial sample order and (2) the algorithm may have a skewed clustering result, if the cluster number estimate is incorrect. Therefore in this section, we will review a couple of adaptive unsupervised learning methods.

Self-organizing map (SOM)

Figure 6.
Self-organizing map of sensor responses on a 5×5 rectangular grid with clusters indicating the sample subgroups labelled using the following scheme: SP-D0, SP-D1, SP¬D2 and SP-D3 are packaged vegetable kept at 10°C for 0–3 days, respectively; EC-D0, EC-D1, EC-D2 and EC-D3 are packaged vegetable inoculated with E. coli on the first day of inoculation and after stored at 10°C for 1–3 days, respectively. (Reprinted with the permission from Elsevier).

SOM is an unsupervised learning neural network. The most significant advantage of SOM is its capability to project a high-dimensional space onto a two-dimensional space while preserving the topology property of high-dimensional data. In terms of structure, SOM is a one-layer neural network. The learning process of SOM is based on a competitive algorithm which can be separated into three major steps: (1) locating the winner neuron whose weight vector is the closest to an input training vector, (2) adjusting the winner weight vector towards the input vector, and (3) updating the weight vectors of the neurons within a neighbouring area of the winner neuron. These three steps reflect the mechanisms of the electric signal transfer among the brain neurons.

SOM was recently applied to detect the pathogen contaminated packaged fresh vegetables (Siripatrawan, 2008). Figure 6 shows a SOM based grouping result for the fresh vegetables with and

without E. coli contamination. SOM has also been used for E-nose based S. aureus infection identification in hospital environment (Dutta et al., 2005).

Adaptive resonance theory based neural network - type 2a (ART-2a)
As an unsupervised learning method, ART-2a is mainly used for cluster analysis. Different from K¬means algorithm where the number of clusters needs to predefined, ART-2a shows its advantageous ability to add a new cluster without disturbing the existing clusters (Carpenter et al., 1991), so it has a potential to be used for real-time sensor data analyses. ART-2a has become the most widely used method for mass spectrometry (MS) data cluster analysis, calibration modelling, and their applications for environmental quality studies (Zhao et al., 2005; Zhao, Hopke, Prather 2008; Song et al., 1999; Phares et al., 2001).
Given the initial cluster centre vectors which usually are randomly selected from the sample set, the learning process of ART-2a can be divided into 4 steps:

1. Randomly select an input sample vector and scale it to unit length.
2. Find the neuron whose cluster vector has the largest resonance to this input vector and call this neuron the "winner". The resonance is estimated by the dot product of the input sample vector and the cluster vector.
3. If the resonance of the winner neuron is larger than a predefined vigilance factor (VF), ρvig, adjust the cluster vector of the winner neuron toward the input sample vector. Otherwise, create a new cluster for this sample vector.
4. Repeat the above steps for all the sample vectors, which is defined as one cycle. Continue this process, until a stopping criterion is reached.

Clearly, the number of clusters is adaptively determined by ART-2a, which makes the cluster analysis more robust and flexible. It can be seen that the vigilance factor is a key parameter to control the cluster number. An overly large vigilance factor could result in an "overly fine" clustering result (the extreme case is one cluster for one mass spectral sample) by generating many homogeneous small clusters, while an overly small vigilance would result in an "overly

coarse" result (Zhao, Hopke, Prather, 2008). In ART-2a, a sample that does not belong to any currently existing clusters is classified into a new cluster. However, during the subsequent training process, this new cluster will only expand but never be merged with its neighbour clusters, even if this new cluster becomes very similar to its neighbour clusters. Therefore, some clusters generated by ART-2a could have a significant overlap among their sample distribution spaces, especially in cases with high vigilance factors. A possible remedy for this problem is to regroup the ART-2a clusters with the same vigilance factor.

Density-based clustering of application with noise (DBSCAN)
DBSCAN presents a very unique clustering process. Different from many other cluster analysis methods including ART-2a, DBSCAN performs a cluster territory expansion process by examining the density and continuity of sample distribution. The entire clustering process is controlled by two parameters: neighbour number (k) and neighbourhood radius (ε). Briefly, the clustering process starts with a randomly selected sample and a cluster is set up for this sample. Controlled by the two predefined parameters, k and ε, the process will expand the territory of this cluster and examine if the cluster can be further expanded. If not, a new cluster will be generated. This searching process will continue until all the samples are clustered. DBSCAN presents a one-step clustering process based on a recursive procedure. The details can be found in the literature (Ester et al., 1996; Daszykowski et al., 2001; 2002).
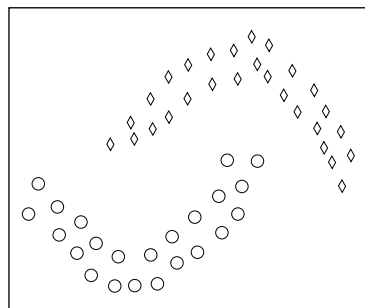
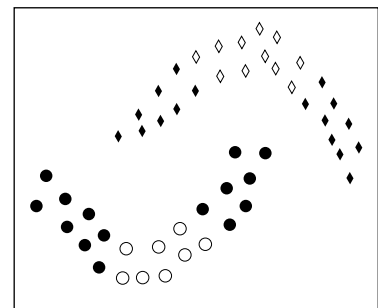DBSCAN   ART-2a

Figure 7.
A schematic illustration of the difference between the clustering principles of DBSCAN and ART-2a.

Figure 7 shows a schematic illustration of the difference between the clustering principles of ART¬2a and DBSCAN. DBSCAN is able

to cluster the samples that have a continuous distribution into one group, while ART-2a and other methods such as K-means divide each group into a couple of small groups based on the sample similarity (distance). With this advantage, DBSCAN surpasses many clustering methods including ART-2a in clustering the data with continuous and irregular distributions.

A comparison study of the clustering effects of DBSCAN and ART-2a on MS data (Zhao, Hopke, Prather, 2008; Zhou et al., 2006) indicates that a proper vigilance factor can produce a reasonable ART-2a clustering result, but an overly fine or "crashed" clustering result for an ART-2a with a high vigilance factor can be recovered by a post-processing strategy. DBSCAN seems to be more effective and robust in the post-processing step than conventional regrouping analysis (Zhao, Hopke, Prather, 2008).

Discussion

After reviewing the methods for each step of a complete machine learning process, this section will discuss some potential problems and feasible suggestions to help ensure machine learning effects. Outlier detection is one of the key issues for data pre-treatment. Because (1) outliers may contain the important information and (2) it can be difficult to detect and confirm outliers in real-time analyses, it is necessary to establish robust models to resist the disturbance of possible outliers. Using weighting functions to adaptively determine the "contribution" of each sample to the modelling effect based on their space or probability distribution is a feasible choice for real-time sensor analyses.

Optimization methods like GA, DE, and SA have been widely used for system feature extraction. As an objective driven searching process, they can adaptively locate different "representative" features or variables for different tasks such as classification or chemical concentration estimation. The key issue in the searching process is to use a proper function (model) to evaluate the fitness and quality of the selected features. It is likely that a set of less "representative" features can also yield a high quality if there is a "super-powerful" model. Therefore, if the feature selection is not only for a high modelling accuracy but also for subsequent mechanism studies such as pathological studies, it

is suggested to employ a relatively straightforward model as a fitness evaluation function to ensure (1) the high fitness is mainly due to the selected features and (2) these features may have sufficient physical meaning.

Another important issue for establishing a reliable supervised learning strategy and preventing over-fitting is to properly make use of the available samples. For a case where there are sufficient samples, a feasible way is to divide the data into three subsets: training, validating, and testing. The training set is used to obtain the model parameters, the validating set is to verify modelling effects and finalize the model construction, and the testing set is used to formally and finally evaluate the modelling effects. For a case with quite limited samples, k-fold cross-validating is good choice for model construction and testing. The extreme case of k-fold cross-validating process is leave-one-out (LOO) strategy, which seems to be the minimum for sample sufficiency requirements. Assuming there are m samples, the LOO method repeats a training-testing process, in which one sample is left out for testing and m-1 samples for training, for m times to cover a small sample set.

Machine learning methods aim to discover the hidden rules that control complex systems. Since the physical laws of these complex systems are not yet clear, a machine learning process based on these statistical and artificial intelligence methods should be a continuously self-adjusting process. In other words, for real applications, we need to be cautious to the results generated by these "grey box" and "black box" models. A validating process and a possible model parameter updating process are always suggested for a reliable model.

Conclusion

This paper has provided a review of the methods that are popularly used in the machine learning step for sensor design. With the increased complexity of sensor systems and their application fields, any of the three steps (data pre-treatment, feature extraction, and system modelling) can be a key factor for a successful machine learning process. Especially, feature extraction and system modelling more and more work together as a system, serving each

other, to obtain a good final result. Selecting "good" features can yield a nice modelling effect and a proper model can help detect "good" features. Integration of linear and nonlinear methods such as RBF-PLS and PCA-BPNN is becoming a popular machine learning method. Gaussian kernel functions are the fundamental element for a group of neural networks including RBFNN, SVM, and PNN. A constructive method to determine the parameters of Gaussian kernel function will lead to an essential improvement of properties of these neural networks. Training time, accuracy, adaptability, stability, and plasticity are the major issues for machine learning processes. There may be a trade-off among them, so integration of the methods with different advantages is a promising direction for the machine learning section of sensor design. Developing transferable machine learning models is another attractive direction for novel sensor design, as it can prevent a complete retraining for a new system or environment. Certainly, instrumentation analysis improvements are a vital force to develop novel machine learning algorithms for sensor design.

Acknowledgements

References

Acevedo F. Javier, Maldonado Saturnino, Domınguez E., Narvaez Arantzazu, Lopez F. (2007), *Probabilistic support vector machines for multi-class alcohol identification*, "Sensors and Actuators B: Chemical", V. 122, n. 1, pp. 227-235

Addison Paul S. (2002), *The illustrated wavelet transform handbook - Introductory theory and applications in science, engineering, medicine, and finance*, London, Institute of Physics Publishing

Aishima Tetsuo (2004), *Correlating sensory attributes to gas chromatography-mass spectrometry profiles and e-nose responses using partial least squares regression analysis*, "Journal of Chromatography A", V. 1054, n. 1-2, pp. 39-46

Alizadeh T., Zeynali Said (2008), *Electronic nose based on the polymer coated SAW sensors array for the warfare agent simulants classification*, "Sensors and Actuators B: Chemical", V. 129, n. 1, pp. 412-423

Ayer Susan, Zhao Weixiang, Davis C. E. (2008), *Differentiation of Proteins, Viruses Using Pyrolysis Gas Chromatography Differential Mobility Spectrometry (PY/GC/DMS) and Pattern Recognition*, "IEEE Sensors Journal", V. 8, n. 9, pp. 1586-1592

Berrueta Luis A., Alonso-Salces Rosa, Heberger Karoly (2007), *Supervised pattern recognition in food analysis*, "Journal of Chromatography A", V. 1158, n. 1-2, pp. 196-214

Bishop Christopher M. (2006), *Pattern recognition and machine learning*, New York, Springer

Branca Andrea, Simonian Paola, Ferrante Marcela, Novas Elena, Negri R. Martin (2003), *Electronic nose based discrimination of a perfumery compound in a fragrance*, "Sensors and Actuators B: Chemical", V. 92, n. 1-2, pp. 222-227

Brezmes Jesús, Cabre Pere, Rojo Sergi, Llobet Eduard, Vilanova Xavier, Correig Xavier (2005), *Discrimination between different samples of olive*

oil using variable selection techniques and modified fuzzy artmap neural networks, "IEEE Sensors Journal", V. 5, pp. 463-470

Buratti Susanna, Benedetti Simona, Scampicchio Matteo, Pangerod Elin C. (2004), *Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue*, "Analytica Chimica Acta", V. 525, pp. 133-139

Cappadona Salvatore, Levander Friedrik, Jansson Maria, James Peter, Cerutti Sergio, Pattini Linda (2008), *Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry*, "Analytical Chemistry", V.80, n.13, pp. 4960-4968

Carpenter Gail A., Grossberg Stephen, Rosen David B. (1991), *ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition*, "Neural Networks", V. 4, pp. 493-504

Chen S., Cowan C. F. N., Grant Peter M. (1991), *Orthogonal least squares learning algorithm for radial basis function networks*, "IEEE Transactions on Neural Networks", V. 2, n. 2, pp. 302-309

Chen Xing, Cao Ming Fu, Li Yi, Hu Wei Jun, Wang Ping, Ying Ke Jing, Pan Hong Ming (2005), *A study of a electronic nose for detection of lung cancer based on a virtual SAW gas sensors array, imaging recognition method*, "Measurement Science & Technology", V. 16, n. 8, pp. 1535-1546

Ciarlini Patrizia, Maniscalco Umberto (2008), *Wavelets, Elman Neural Networks for monitoring environmental variables*, "Journal of Computational and Applied Mathematics", V. 221, n. 2, pp. 302-309

Comon Pierre (1994), *Independent Component Analysis, a New Concept*, "Signal Processing", V. 36, n. 3, pp. 287-314

Crowe Cameron M. (1989), *Test of Maximum Power for Detection of Gross Errors in Process Constraints*, "Aiche Journal", V. 35, n. 5, pp. 869-872

Cybenko G. (1989), *Approximation by superposition of a sigmoidal function*, "Mathematics of Control, Signals and Systems.", V. 2, n. 4, pp. 303-314

Daszykowski Michael, Walczak Beata, Massart Desire L. (2002), *Representative subset selection*, "Analytica Chimica Acta", V. 468, n. 1, pp. 91-103

Daszykowski Michael, Walczak Beata, Massart Desire L. (2001), *Looking for natural patterns in data. Part 1. Density-based approach*, "Chemometrics and Intelligent Laboratory Systems", V. 56, n. 2, pp. 83-92

Daubechies Ingrid (1988), *Orthonormal bases of compactly supported wavelets*, "Communications on Pure Applied Mathematics", V. 41, n. 4, pp. 909-996

Di Natale Corrado, Martinelli Eugenio, D'Amico Arnaldo (2002), *Counteraction of environmental disturbances of electronic nose data by independent component analysis*, "Sensors and Actuators B: Chemical", V. 82, n. 2-3, pp. 158-165

Dodd T. H., Hale S. A., Blanchard S. M. (2004), *Electronic nose analysis of Tilapia storage*, "Transactions of the ASAE", V. 47, n.1, pp. 135-140

Dutta Ritabrata, Dutta Ritabran (2006), *"Maximum probability rule" based classification of MRSA infections in hospital environment: Using electronic nose*, "Sensors and Actuators B: Chemical", V. 120, n. 1, pp. 156-165

Dutta Ritaban, Das Aruneema, Stocks Nigel, Morgan David (2006), *Stochastic resonance-based electronic nose: A novel way to classify bacteria*, "Sensors and Actuators B: Chemical", V. 115, n. 1, pp. 17-27

Dutta Ritaban, Morgan David, Baker Nicky, Gardner Julian, Hines Evor (2005), *Identification of Staphylococcus aureus infections in hospital environment: electronic nose based approach*, "Sensors and Actuators B: Chemical", V. 109, n. 2, pp. 355-362

Esposito Anna, D'Andria Pierluigi (2003), *An adaptive learning algorithm for ECG noise and baseline drift removal*, in *Neural Nets*, V. 2859, Heidelberg, Springer Berlin, pp. 139-147

Ester Martin, Kriegel Hans-Peter, Sander Jörg, Xu Xiaowei (1996), *A densitybased algorithm for discovering clusters in large spatial databases with noise*. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining KDD-96, 2-4 August, Portland, Oregon

Fisher Ronald (1936), *The Use of Multiple Measurements in Taxonomic Problems*, "Annals of Eugenics", V. 7, n. 2, pp. 79-188

Gardner Julian, Shin Hyun Woo, Hines Evor (2000), *An electronic nose system to diagnose illness*, "Sensors and Actuators B: Chemical", V. 70, n. 1-3, pp. 19-24

Goldberg David E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Boston, MA, Kluwer Academic Publishers

Hibbert D. Brynn (1993), *Genetic algorithms in chemistry*, "Chemometrics and Intelligent Laboratory Systems", V. 19, n. 3, pp. 277-293

Hoskuldsson Agnar (1988), *PLS Regression Methods*, "Journal of Chemometrics", V. 2, pp. 211-228

Hush D. R., Horne B. G. (1993), *Progress in supervised neural networks*, "IEEE Signal Processing Magazine", V. 10, n. 1, pp. 8-39

Kecman Vojislav (2001), *Learning and Soft Computing, Support Vector Machines, Neural Networks and Fuzzy Logic Models*, Cambridge, The MIT Press, MA, USA

Kim Namyong, Byun Hyung-Gi, Persaud Krishna C. (2007), *Normalization approach to the stochastic gradient radial basis function network algorithm for odor sensing systems*, "Sensors and Actuators B: Chemical", V. 124, n. 2, pp. 407-412

Kirkpatrick S., Gelatt C. D., Vecchi M. P. (1983), *Optimization by Simulated Annealing*, "Science", V. 220, n. 4598, pp. 671-680

Krebs Melissa D., Tingley Robert D., Zeskind Julie E., Holmboe Maria E., Kang Joung-Mo, Davis Cristina E. (2006), *Alignment of gas chromatography-mass spectrometry data by landmark selection from complex chemical mixtures*, "Chemometrics and Intelligent Laboratory Systems", V. 81, n. 1, pp. 74-81

Krebs Melissa D., Kang Joung-Mo, Cohen Sarah J., Lozow Jeffrey B., Tingley Robert D., Davis Cristina E. (2006), *Twodimensional alignment of differential mobility spectrometer data*, "Sensors and Actuators B: Chemical", V. 119, n. 2, pp. 475-482

Krebs Melissa D., Mansfield Brian, Yip Ping, Cohen Sarah J., Sonenshein Abraham L., Hitt Ben A., Davis Cristina E. (2006), *Novel technology for rapid species-specific detection of Bacillus spores*, "Biomolecular Engineering", V. 23, n. 2-3, pp. 119-127

Krebs Melissa D., Tingley Robert D., Zeskind Julie E., Kang Joung-Mo, Holmboe Maria E., Davis Cristina E. (2005), *Autoregressive modeling of analytical sensor data can yield classifiers in the predictor coefficient parameter space*, "Bioinformatics", V. 21, n. 8, pp. 1325-1331

Krier Catherine, Rossi Fabrice, Francois Damien, Verleysen Michel (2008), *A data-driven functional projection approach for the selection of feature ranges in spectra with ICA or cluster analysis*, "Chemometrics and Intelligent Laboratory Systems", V. 91, pp. 43-53

Kuske Martyna, Romain Anne-Claude, Nicolas Jacques (2005), *Microbial volatile organic compounds as indicators of fungi. Can an electronic nose detect fungi in indoor environments?*, "Building and Environment", V. 40, n. 6, pp. 824-831

Li Changying, Heinemann Paul (2007), *A comparative study of three evolutionary algorithms for surface acoustic wave sensor wavelength*

selection, "Sensors and Actuators B: Chemical", V. 125, n.1, pp. 311-320

Llobet Eduard, Gualdron O., Vinaixa Maria, El-Barbri Nourredine, Brezmes Jesus, Vilanova X., Bouchikhi Benachir, Gomez R., Carrasco J. A., Correig Xavier (2007), *Efficient feature selection for mass spectrometry based electronic nose applications*, "Chemometrics and Intelligent Laboratory Systems", V. 85, n. 2, pp. 253-261

Mah Richard S. H., Tamhane Ajit C. (1982), *Detection of Gross Errors in Process Data*, "Aiche Journal", V. 28, n. 5, pp. 828-830

Mallet Stephan G. (1989), *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, "IEEE Transaction on Pattern Analysis and Machine Intelligence", V. 11, n. 7, pp. 674-693

Marin Sonia, Vinaixa Maria, Brezmes Jesús, Llobet Eduard, Vilanova Xavier, Correig Xavier, Ramos Antonio J., Sanchis Vicent (2007), *Use of a MS-electronic nose for prediction of early fungal spoilage of bakery Products*, "International Journal of Food Microbiology", V. 114, pp. 10-16

Markovic Milan Z., Milosavljevic Milan M., Kovacevic Blanko D. (2002), *Quadratic classifier with sliding training data set in robust recursive AR speech analysis*, "Speech Communication", V. 37, n. 3-4, pp. 283-302

Michalewicz Z., Janikow C. Z., Krawczyk J. B. (1992), *A Modified Genetic Algorithm for Optimal-Control Problems*, "Computers & Mathematics with Applications", V. 23, n. 12, pp. 83-94

Monge Maria Eugenia, Bulone Donatella, Giacomazza Daniela, Bernik Delia, Negri R. Martin (2004), *Detection of flavour release from pectin gels using electronic noses*, "Sensors and Actuators B: Chemical", V. 101, n. 1-2, pp. 28-38

Moreno-Baron Laura, Cartas Raul, Merkoci Arben, Alegret Salvador, del Valle Manuel, Leija Lorenzo, Hernandez Pablo R., Munoz Roberto (2006), *Application of the wavelet transform coupled with artificial neural networks for quantification purposes in a voltammetric electronic tongue*, "Sensors and Actuators B: Chemical", V. 113, n. 1, pp. 487-499

Munoz Alberto, Muruzabal Jorge (1998), *Self-organizing maps for outlier detection*, "Neurocomputing", V. 18, n. 1-3, pp. 33-60

Paatero Pentti (1997), *Least squares formulation of robust non-negative factor analysis*, "Chemometrics and Intelligent Laboratory Systems", V. 37, n. 1, pp. 23-35

Pearson G. A. (1977), *General Baseline-Recognition and Baseline-Flattening Algorithm*, "Journal of Magnetic Resonance", V. 27, pp. 265-272

Phaisangittisagul Ekachai, Nagle H. Troy (2008), *Sensor selection for machine olfaction based on transient feature extraction*, "IEEE Transactions on Instrumentation and Measurement", V. 57, n. 2, pp. 369-378

Phares Denis J., Rhoads Kevin P., Wexler Anthony S., Kane David B., Johnston M. V.(2001), *Application of the ART-2a Algorithm to Laser Ablation Aerosol Mass Spectrometry of Particle Standards*, "Analytical chemistry", V. 73, n. 10, pp. 2338-2344

Prescott P. (1975), *Approximate Test for Outliers in Linear-Models*, "Technometrics", V. 17, pp. 129-132

Qian Tao, Li Xiaokun, Ayhan Bulent, Xu Roger, Kwan Chiman, Griffin Tim (2006), *Application of Support Vector Machines to Vapor Detection and Classification for Environmental Monitoring of Spacecraft*. Proceedings ISNN 2006 "Lecture notes in computer science" International Symposium on Neural Networks, Chengdu, China, May 28-June 1, 2006

Roncaglia Alberto, Elmi Ivan, Dori Leonello, Rudan Massimo (2004), *Adaptive K-NN for the detection of air pollutants with a sensor array*, "IEEE Sensors Journal", V. 4, n. 2, pp. 248-256

Scorsone Emmanuel, Pisanelli Anna Maria, Persaud Krishna C. (2006), *Development of an electronic nose for fire Detection*, "Sensors and Actuators B: Chemical", V. 116, pp. 55-61

Shnayderman Marianna, Mansfield Brian, Yip Ping, Clark Heather A., Krebs Melissa D., Cohen Sarah J., Zeskind Julie E., Ryan Edward T., Dorkin Henry L., Callahan Michael V., Stair Thomas O., Gelfand Jeffrey A., Gill Christopher J., Hitt, Ben A., Davis Cristina E. (2005), *Species-specific bacteria identification using differential mobility spectrometry and bioinformatics pattern recognition*, "Analytical Chemistry", V. 77, n. 18, pp. 5930-5937

Shusterman Vladimir, Shah Syed Ismail, Beigel Anna, Anderson Kelly P. (2000), *Enhancing the precision of ECG baseline correction: selective filtering and removal of residual error*, "Computers and Biomedical Research", V. 33, n.2, pp. 144-160

Siripatrawan Ubonrat (2008), *Self-organizing algorithm for classification of packaged fresh vegetable potentially contaminated with foodborne pathogens*, "Sensors and Actuators B: Chemical", V. 128, n. 2, pp. 435-441

Smidl Vaclav, Quinn Anthony (2005), *Mixture-based extension of the AR model and its recursive Bayesian identification*, "IEEE Transactions on Signal Processing", V. 53, n. 9, pp. 3530-3542

Song Xin-Hua, Hopke Philip K., Fergenson David P., Prather Kimberly A. (1999), *Classification of single particles analyzed by ATOFMS using an artificial neural network*, ART-2A, "Analytical chemistry", V. 71, n. 4, pp. 860-865

Storn Rainer, Price Kenneth (1995), *Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces*, Technical Report TR95-012, ICSI International Computer Science Institute, http://www.icsi.berkeley.edu/cgi-bin/pubs/publication.pl?ID=000952

Tchoupo Guy Narcisse, Guiseppi-Elie Anthony (2005), *On pattern recognition dependency of desorption heat, activation energy, and temperature of polymer-based VOC sensors for the electronic NOSE*, "Sensors and Actuators B: Chemical", V. 110, n. 1-2, pp. 81-88

Trihaas Jeorgos, Vognsen Lene, Nielsen Per V. (2005), *Electronic nose: New tool in modeling the ripening of Danish blue cheese*, "International Dairy Journal", V. 15, pp. 679-691

Ubeyli Elif Derya, Guler Inan (2004), *Spectral analysis of internal carotid arterial*. Doppler signals using FFT, AR, MA, and ARMA methods, "Computers in Biology and Medicine", V. 34, n. 4, pp. 293-306

Vestergaard Jannie, Haugen John-Eric, Byrne Derek (2006)(a), *Application of an electronic nose for measurements of boar taint in entire male pigs*, "Meat Science", V. 74, n. 3, pp. 564-577

Vinaixa Maria, Marin Sonia, Brezmes Jesús, Llobet Eduard, Vilanova Xavier, Correig Xavier, Ramos Antonio, Sanchis Vicent (2004), *Early detection of fungal growth in bakery products by use of an electronic nose based on mass spectrometry*, "Journal of Agricultural and Food Chemistry", V. 52, n. 20, pp. 6068-6074

Yadava R. D. S., Chaudhary Ruchi (2006), *Solvation, transduction and independent component analysis for pattern recognition in SAW electronic nose*, "Sensors and Actuators B: Chemical", V. 113, n. 1, pp. 1-21

Yan Xuefeng, Zhao Weixiang (2009), *4-CBA concentration soft sensor based on modified back propagation algorithm embedded with ridge regression*, "Intelligent Automation and Soft Computing", V. 15, n.1, pp. 41-51

Yu Hui Chun, Wang Jun, Zhang Hong Mei, Yu Yong, Yao Cong (2008), *Identification of green tea grade using different feature of response signal from E-nose sensors*, "Sensors and Actuators B: Chemical", V. 128, n. 2, pp. 455-461

Walczak Beata, Massart Desire L. (1996), *The Radial Basis Functions - Partial Least Squares approach as a flexible non-linear regression technique*, "Analytica Chimica Acta", V. 331, n. 3, pp. 177-185

Whitehead B. A., Choate T. D. (1996), *Cooperative-Competitive genetic evolution of radial basis function centers and widths for time series prediction*, "IEEE Transactions on Neural Networks", V. 7, n. 4, pp. 869-880

Wold Svante, Esbensen Kim, Geladi Paul (1987), *Principal Component*

Analysis, "Chemometrics and Intelligent Laboratory Systems", V. 2, pp. 37-52

Zhao Weixiang, Morgan Joshua, Davis Cristina (2008), *Gas Chromatography Data Classification Based on Complex Coefficients of an Autoregressive Model*, "Journal of Sensors", doi:10.1155/2008/262501

Zhao Weixiang, Hopke Philip K., Prather Kimberly A. (2008), *Comparison of two cluster analysis methods using single particle mass spectra*, "Atmospheric Environment", V. 42, n. 5, pp. 881-892

Zhao Weixiang, Hopke Philip K. (2006), *Source identification for fine aerosols in Mammoth Cave National Park*, "Atmospheric Research", V. 80, n. 4, pp. 309-322

Zhao Weixiang, Hopke Philip, Qin Xueying, Prather Kimberly (2005), *Predicting bulk ambient aerosol compositions from ATOFMS data with ART-2a and multivariate analysis*, "Analytica Chimica Acta", V. 549, n. 1-2, pp. 179-187

Zhao Weixiang, Hopke Philip K. (2004), *Source apportionment for ambient particles in the San Gorgonio wilderness*, "Atmospheric Environment", V. 38, n. 35, pp. 5901-5910

Zhao Weixiang, Chen Dezhao, Hu Shangxu (2004), *Detection of outlier and a robust BP algorithm against outlier*, "Computers & Chemical Engineering", V. 28, n. 5, pp. 1403-1408

Zhou Liming, Hopke Philip K., Venkatachari Prasanna (2006), *Cluster analysis of single particle mass spectra measured at Flushing*, NY, "Analytica Chimica Acta", V. 555, n. 1, pp. 47-56

Sintesi

Analisi dello stato dell'arte nella progettazione e uso dei sensori secondo gli schemi del machine learning.

Il concetto di machine learning nasce nell'ambito dell'intelligenza artificiale, ramo dell'informatica (intesa qui nell'accezione più ampia di disciplina dedita sia alla progettazione che alla realizzazione delle componenti hardware e software) il cui fine è la realizzazione di calcolatori capaci di svolgere i ragionamenti tipici della mente umana. All'interno dell'intelligenza artificiale, il machine learning - o approfondimento automatico - si occupa della realizzazione di sistemi che, partendo dall'osservazione di dati, permettono la sintesi di nuove conoscenze.

Nel corso del tempo l'idea sottesa a tale tipologia di sistemi, ovvero i compiti assegnati alle due componenti hardware e software, si è naturalmente modificata. Inizialmente, infatti, la sintesi era lasciata ad algoritmi dedicati i quali elaboravano la totalità dei dati; la componente hardware, pur se differente a seconda dei casi, non svolgeva altra attività se non quella di rilevazione delle informazioni (dati). Le due componenti hardware e software avevano quindi domini e funzioni disgiunte e complementari: la prima dedicata alla sola acquisizione e quantificazione delle grandezze sotto studio; la seconda efficace nella determinazione delle acquisizioni utili, e nella sintesi e verifica delle nuove descrizioni della realtà. Tuttavia, i progressi tecnologici nel tempo hanno portato questa differenza ad essere meno netta, in particolare il ruolo dell'hardware si è ampliato portando alla progettazione e realizzazione di sensori capaci di un pretrattamento dei dati, diminuendo così il carico della componente algoritmica e, di conseguenza, velocizzandone la sintesi.

Un esempio noto che può illustrare il passaggio da un hardware passivo ad uno attivo è dato dai computer e il gioco degli scacchi. È indubbio infatti che un qualsiasi elaboratore, una volta corredato del software adatto, possa giocare a scacchi, ma sono quelli dedicati - quelli ovvero in cui anche la componentistica è progettata appositamente per tale gioco, quali Deep Blue - a sintetizzare le strategie migliori.

Il lavoro dell'equipe composta da Bhushan, Davis, Santamaria, Simon e Zhao illustra l'attuale schematizzazione di un processo costruito tramite le tecniche dell'apprendimento automatico, suddividendolo nelle tre fasi fondamentali: pretrattamento dei dati, ridimensionamento e selezione, e modellizzazione. Di particolare interesse è la descrizione dagli algoritmi maggiormente efficaci per ognuna di queste fasi e le interessanti considerazioni per migliorare la componente hardware che da queste prendono spunto. Questi possibili perfezionamenti hanno alla base l'applicazione dell'idea stessa di machine learning alle unità destinate alla rilevazione dei dati: i sensori. Con altre parole, il sensori (più correttamente, i sistemi sensore) che prima si dedicavano soltanto all'acquisizione dei dati, vengono ora considerati strutture dotate di capacità di apprendimento le quali, analizzando le informazioni

da se stesse ricavate, formulano delle regole per la discriminazione dei dati utili, per il calcolo della numerosità delle informazioni necessarie da acquisire e, soprattutto, possono effettuare test atti a verificare se le regole ricavate siano "buone" e, quindi, se continuare ad applicarle o cercarne altre. È nella deduzione e verifica delle regole che si manifesta l'apprendimento automatico: il sensore impara dalla sue stesse misurazioni ad acquisire un maggior numero di dati utili, da passare successivamente agli algoritmi di analisi, e, insieme, verifica l'affidabilità delle proprie deduzioni.

In questo è evidente la somiglianza con parte della metodologia scientifica umana: osservazione, formulazione, verifica.

Chiaramente, per quanto possano essere precise e a prova di errore le migliorie apportate al sistema sensori, questi devono comunque essere soggetti ad un controllo che ne verifichi a sua volta il corretto funzionamento evitando quindi una loro trasformazione in scatola nera. Ciononostante, considerato l'impiego di queste tecnologie nei campi più disparati (dalle diagnosi mediche ai monitoraggi ambientali, dal controllo della qualità dei generi alimentari all'analisi dei processi industriali), il loro miglioramento non può che avere una ricaduta positiva (e, forse, anche inaspettatamente vasta) nella qualità della vita.