International Journal of Information Technology and Knowledge Management
January June 2009, Volume 2, No. 1, pp. 21-24

PATTERN RECOGNITION SYSTEM FOR E-NOSE

Tapsi Garg*

INTRODUCTION

An odour is the description of the sensation perceived when volatile chemicals are sensed by the biological nose. The odours that we perceive in the world around us arise through the interaction of molecules with sensory cells located in our nose. Typically, the odour of a flower or of a foodstuff or other natural material consists of a complex mixture of molecules present in different concentrations. In other cases a single, pure chemical compound can be the dominant odour. The human nose is widely used as an analytical tool in industry today, e.g. to assess the quality of food-stuff, drinks, perfumes and many other household products. Panels are setup at sites manufacturing products with odour specifications. Despite having high performance at low concentrations the human nose cannot be used economically, because these panels are expensive to train and maintain and give subjective assessments which can be adversely affected by external parameters such as illness and fatigue. They are also unsuited for use in aggressive environments and with toxic and obnoxious odours. Furthermore, it is difficult to replicate an olfactory panel at different geographical sites. Consequently, there is considerable need for an instrument that could mimic the human sense of smell and be used in industrial applications. Gas chromatography-mass spectroscopy (GC-MS) is one of the most widely used techniques for the analysis of complex mixtures to be separated and the individual constituents identified and quantified. As the technique became widespread and more sophisticated, it was possible to separate and chemically identify the dozens or hundreds of individual substances present in food, flavour and fragrance products. However, this is a complex, expensive and time-consuming task which requires a well-equipped analytical laboratory and skilled staff. Hence, there is enormous demand for an electronic instrument that can mimic the human sense of smell and provide low-cost and rapid sensory. One of the pioneering steps, towards the concept of an electronic nose as an intelligent chemical array sensor system for odour classification, was made in 1982 by Persaud and Dodd (Persaud et al. 1982).

* Sr. Lecturer (CS), RKGIT, Ghaziabad, E-mail: *tapsig@gmail.com*

Electronic Nose

According to Gardner (Gardner *et al*. 1999) "*An electronic nose is an instrument, which comprises an array of electronic chemical sensors with partial specificity and an appropriate pattern recognition system, capable of recognising simple or complex odours"*

Two Main Components

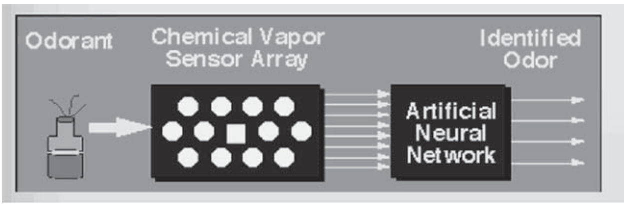

Figure 1: Block Diagram of E-Nose

- Sensing System (Array of chemical sensors, single sensing device)
- Automated Pattern Recognition System (Statistical, ANN approaches)

Since the mid 1980s there has been increasing interest in the development of so-called 'electronic noses', i.e., an electronic instrument that is capable of detecting and recognising complex odours. They are based around an array of sensors, each of which has a partial specificity. This array produces a finger print of the odour which is used by an appropriate pattern recognition system to identify the odour through comparison with a reference library of previously obtained measurements of known samples. A chemical sensor consists of a chemical sensitive layer and a transducer.

The chemical sensitive material captures the interaction with the analyte molecules present in the environment and generates a physical change which is sensed by the transducer that converts the signal into an analogue electrical output. The transduction mechanism makes available several physical signals where it has been widely used the approach of electrical measurements (current, resistance, voltage, and capacitance), mass changes, heat generation, and measurements of optical changes (absorption, fluorescence and reflectivity).

A most important consideration for the successful operation of an odour sensor is its long-term stability. *Drift* is defined as "*a gradual change in any quantitative characteristic that is supposed to remain constant*". Thus a drifting chemical sensor does not give exactly the same response even if it is exposed to exactly the same environment for a long time.

The main aim of the research work is to investigate optimum pattern recognition approaches for electronic nose systems over long period – where both sensors and analytes may be drifting. Although chemical patterns from the sensor array should be the same for a particular sample, the actual responses are affected by many factors such as temperature, humidity and sensor drift. If no drift correction of the sensor signals is made, the model will have a continuous need for re-calibration. Hence, a new architecture called mSom (Distante *et al*. 2002) (where *m* is the number of independent maps), has been used on the basis of the self-organising map (SOM) theory developed by Kohonen (Kohonen 1982) Unlike other neural techniques where the input-output learned mapping is static, the architecture here presented reduces the drift problem (thus increasing the time between re-calibrations) by addressing the problem of dynamic input-output mapping.

PATTERN RECOGNITION ALGORITHMS

Pattern recognition is defined as the process of classification of data by comparison to known patterns. Patterns are typically described in terms of multidimensional data vectors, where each component is called a *feature*. The aim of a pattern recognition system is to associate each pattern with one of a number of possible *pattern classes* (or simply *classes*). Obviously, different patterns should be associated with the same class or with different classes depending on whether they are characterised by similar or dissimilar features, respectively. In the case of the electronic nose, the patterns and the classes are, respectively, the responses of the sensor array to odorants and the odorants being considered. In order to develop a pattern recognition system, the sample data are split into two sets, namely, the *training set* and the *test set*. The training set is used to establish the design parameters of the pattern recognition system, whereas the test set contributes to evaluate the system performance. Typically, the performance of the pattern recognition system is measured by computing the percentage of correctly recognised patterns on all the patterns presented to the system. Of course, the performance of the pattern recognition system should be as independent as possible from how the sample data are split into training and test sets. Several different data processing and pattern recognition techniques have been used in the literature to recognise signals produced by sensor arrays (Di Natale *et al*. 1995a). These include linear pattern recognition techniques, such as principal component analysis and cluster analysis (Gardner 1991) and non-linear pattern recognition techniques, such as classical multivariate analysis and artificial neural network (ANN) algorithms (Di Natale *et al*. 1995b). As the relationship between the signal produced by a sensor and an odorant concentration is usually non-linear, non-linear pattern recognition techniques are generally more successful than linear ones. However, the success of each technique heavily depends on the preliminary selection of the features, which are used in the recognition process.

The algorithmic part of an odour discrimination system consists of three steps:

(i) Signal conditioning and feature extraction

(ii) Dimensionality reduction and

(iii) Classification.

The role of the first step is to segment the pattern of interest from the sensor response, remove noise, normalise the pattern and any other operation that contributes in defining a compact representation of the pattern. Feature reduction should provide a small number of informative features in order to make the learning task simpler. Classification tasks address the problem of identifying an unknown sample as one from a set of recognisable gases.

Neural Network Methods

Adaptive resonance theory (ART) classifier has been used (Shukla *et al*. 1998) for odour sensing using four SnO_2 sensors to sense four gases (acetone, ethyl methyl ketone, carbon tetrachloride and xylene) because of its unsupervised and self-organising nature. It avoids the drawback associated with static feedforward neural networks trained with locally optimal backpropagation-type training algorithms and removes all the disadvantages of feedforward ANN e.g. local minima, slow convergence, non-linear activation function, random weight initialisation etc.

A new drift counteraction method based on the on-line learning ability of a fuzzy ARTMAP neural network was presented in (Paniagua *et al*. 2003). The method has been demonstrated in the identification of three different vapours ethanol, ammonia and benzene using commercially available 12-element tin oxide gas sensor array in the presence of an artificially induced nonlinear drift. To induce artificial drift in dataset, the heating voltage of the sensors was linearly increased in six steps. Four replicate measurements of each species for every heating voltage were performed over a period of six weeks giving a total of 72 measurements in testing dataset. It used calibration measurements performed at a fixed rate to retrain the neural network model. Different neural networks like MLP, LVQ and fuzzy ARTMAP were tested for there drift counteraction capability. Networks were trained using training dataset (free from drift) and then retrained using calibration measurements selected at random (one calibration measurement per species and heating voltage). The performance of the retrained network was evaluated over the remaining drifted measurements (i.e. those not selected for retraining). No matter which neural network is used the drift compensation method restores the discrimination ability of the system (i.e. vapour recognition rate is 100% over the 10 trials). However, the fuzzy ARTMAP network is preferred since it is trained in just two epochs and this compares very favourably with the MLP and LVQ. Additionally, while the fuzzy ARTMAP can perform fast on-line learning, the MLP and LVQ have to be trained off-line. The identification rate rises from 61.5% to 100% after drift compensation using the fuzzy ARTMAP and would allow for an automatic recalibration of the PARC engine in EN and gas analysers.

The radial basis function (RBF) network is an artificial neural network with good generalisation ability that trains rapidly, while exhibiting none of back-propagation's training pathologies such as local minima problems. The performance of RBF as a classifier is highly dependent on the centers and widths in basis function (usually a Gaussian response function) used in the hidden layer. Most of the learning algorithms for RBF networks are divided into the two stage processing. Firstly, a clustering algorithm (e.g. *k*-means, fuzzy *c*-means or genetic fuzzy *c*-means) is applied to the input patterns in order to determine the centers for hidden layer units. After the centers are fixed, the widths are determined in a way that reflects the distribution of the centers and input patterns. Once the centers and weights are fixed, the weights between hidden and output layer are trained by a single shot training using singular value decomposition (i.e. weights were initialised using singular value components) or by the least-means-squares (LMS) algorithm. This two-stage method provides some useful solutions in pattern classification problem.

The Potential Function (PF) method is a well-known classification method employed in supervised pattern recognition. In the PF classification method, each pattern of that set is considered as a charge surrounded by a short-range potential field in the pattern space, say a Gaussian potential, for instance. Given a class, its cumulative potential field is obtained by adding up the individual potentials of its known members, i.e., its patterns included in the calibration set. Afterwards a test pattern is classified into the class that has the highest cumulative potential at the point where the test pattern lies. (Davide et al. 1994) proposed a new classification method, which generalises the potential function (PF) method to neural implementation in an unsupervised environment. It used SOM for time dependent analysis and drift effects for two-odour recognition experiment. PF method has been adapted to an unsupervised environment, where the definition of the classes is not previously provided and no training set is available. In case of drift, the network has proven to be successful in counteracting any variation of input statistics by continuous adjustment of neuron weights, i.e., moving the neuron images in the pattern space. Network was able to reject the noise and other disturbances as well.

SOMs show a particular ability for solving the problem of classification in pattern recognition. Unlike classical statistical methods, an SOM does not require any preventive knowledge about the statistical distribution of the patterns in the environment. SOMs combine competitive learning with dimensionality reduction by smoothing the clusters with respect to an a priori grid of neurons. Among the other properties of the SOMs, one of the most amazing is their adaptability. This allows the SOM, under certain condition to adapt itself to recover from eventual change of the class properties of the environment (such as their numbers or their statistical characteristics). The long-term drifts produce dispersion in the patterns and it can eventually change the cluster distribution in the data space, making useless the internal representation reached by the SOM netwok during the training phase.

Objectives of Research

Gas-sensor array technology combined with various pattern recognition methods is widely used in the gas analysis field. The aim of the development of electronic nose technology is to provide cheap and small online instruments for fast discrimination, recognition and quantification of specific chemicals, odours, or toxic substances (Gopel 1998) with high spatial and time resolution. The pattern recognition system is trained on known patterns collected using the data acquisition electronics/software. These patterns form the knowledge base of the system. The mapping of each gas concentration or class identification is made using these trained patterns during the later measurements. But sensor drift will eventually destroy the initially trained pattern recognition capability, so that it is very difficult for the system to classify or predict the exact concentrations of the gas monitored. There are many factors such as poisoning, ageing or environmental changes that can cause the sensor drift.

The scope of the research is to develop a dynamic pattern recognition system to treat the generic drift problem of array based sensors. Our aim is to develop a pattern recognition software that is responsible for interpreting the output from the gas sensors in the real environment where both the sensors and the environment are likely to drift. The software required must be fast and reliable at distinguishing gases and / or odours of interest as well as be able to compensate for the drift in the sensor array data over a long run. A pattern recognition system will be developed in Visual C++/C# using the multiple Self Organising Maps (mSom) that will be able to adapt itself to the changing input probability distribution because of the sensor drift and can recognise the patterns in the real environment conditions over a long period of time.

In the design of electronic noses, as with other types of transducer, attention needs to be paid to the requirements of the user and in what areas the instrument is likely to be applied. For such an instrument to be useful in a variety of situations, it needs to be flexible enough to adapt to changing requirements of odour discrimination. In view of the array nature of the human nose, it makes sense to design an instrument based on arrays of elements and to use classifier software that would be able to counteract the slow systematic drift problem of the sensors when used for a long period.

References

[1] Christopher J. C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition,* Bell Laboratories, *Lucent Technologies*, 121–162.

[2] Distante, C., Siciliano, P., & Persaud, K. C. “Dynamic Cluster Recognition with Multiple Self-Organising Maps”, *Pattern Analysis & Applications*, **5**, (3), (2002), 306–315.

[3] Gardner, J. W. & Bartlett, P. N. *Electronic Noses: Principles and Applications* Oxford University Press, Oxford, (1999).

[4] Gopel, W. “Chemical Imaging: I. Concepts and Visions for Electronic and Bioelectronic Noses”, *Sensors and Actuators B: Chemical*, **52**, (1-2), (1998), 125–142.

[5] Persaud, K. C. & Dodd, C. “Analysis of Discrimination Mechanisms in the Mammalian Olfactory System using a Modal Nose”, *Nature*, **5881**, (299), (1982), 352–355.

[6] Sensors and Actuators B: Chemical **116**[1-2], (2006), 1–226.

[7] Simon Hykin, P. N., Neural Networks: A Comprehensive Foundation, (2006), 340–348

[8] Strike, D. J., Meijerink, M. G. H., & Koudelka-Hep, M. “Electronic Noses - A Mini-Review”, Fresenius’ *Journal of Analytical Chemistry*, **364**, (6), (1999), 499–505.