IMTC 2004 - Instrumentation and Measurement
Technology Conference
Como, Italy, May 18–20, 2004

Gas Identification Algorithms for Microelectronic Gas Sensor

S. Brahim Belhouari, A. Bermak, G. Wei, P. C. H. Chan

Hong Kong University of Science and Technology,
EEE Department, Clear Water Bay, Kowloon, Hong Kong.
Phone: +852 2358 8844, Fax: +852 2358 1485,
E-mail: eebelhou@ust.hk, eebermak@ust.hk, eegfwei@ust.hk, eepchan@ust.hk.

Abstract *– Gas identification represents a big challenge for pattern recognition systems due to several particular problems. The aim of this study is to compare the accuracy of a range of advanced and classical pattern recognition algorithms for gas identification from sensor array signals. Density estimation is applied in the construction of classifiers through the use of Bayes rule. Experiments on real sensors' data proved the effectiveness of the approach with an excellent classification performance. We compare the classification accuracy of different density models with several neural networks architectures. On our gas sensors data, the best performance was achieved by Gaussian mixture models with more than 92% accuracy.*

Keywords *– Classification, gas sensor array, density models, neural networks, pattern recognition.*

I. INTRODUCTION

During the last decade, increasing attention has been given to the development of Microelectronic gas sensors. Among various types of sensors, the micro-hotplate based SnO_2 thin film sensors offer a number of interesting features and are particularly attractive for their practical interest [1]. Indeed, these devices feature high sensitivity, lower power consumption, as well as compactness and compatibility with semiconductor technology. Unfortunately, thin film sensors, (as do all gas sensors) suffer from a number of shortcomings such as non selectivity and nonlinearities of the sensor's response. Pattern recognition algorithms combined with a gas sensor array have been traditionally used to address these issues [2]. In fact, a gas sensor array permits to improve the selectivity of the single gas sensor, and shows the ability to classify different odors. An array of different gas sensors is used to generate a unique signature for each odor. After a preprocessing stage, the resulting feature vector is used to solve a given classification problem, which consists of identifying an unknown sample as one from a set of previously learned gases. Significant work has been devoted to design a successful pattern analysis system for machine olfaction [2]. Various kinds of flexible pattern recognition algorithms have been used for classifying chemical sensor data. Most notably neural networks have been exploited, in particular multilayer perceptrons (MLP), radial basis functions (RBF) and self organizing maps (SOM) [2]. Other methods based on the class-conditional density estimation have been used, such as quadratic and K nearest neighbors (KNN) classifiers.

Recently, a new family of semiparametric methods based on mixture distributions, have been successfully applied for a number of applications such as speech recognition [3] and image retrieval [4]. Despite their great potential as classifiers, density model-based approaches have not been exploited for machine olfaction and electronic nose applications. In this paper, we will present a gas classification approach based on either class-conditional density estimation using different density models or discriminant functions. The ability of the proposed models to perform gas identification will be compared using an experimentally obtained dataset. An integrated sensor array has been designed with the aim of the identification of combustion gases. Data collected from the microelectronic gas sensor constituted of a VLSI chip including 8 sensors would first undergo a preprocessing stage before being fed to the classifier. A total of six classifiers operating on PCA, Neuroscale and LDA projections will be compared.

II. DATA DESCRIPTION

Measurements have been done using an experimental setup consisting of a special sensor chamber equipped with gas pumps and mass flow controllers as well as a data acquisition board (Figure 1).

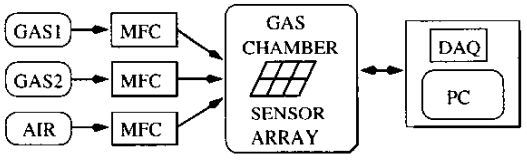

Fig. 1. Scheme of the experimental setup. MFC stands for Mass Flow Controller.

The sensor array composed of 8 micro-hotplate based SnO_2 thin film gas sensors, have been used. Four sensors with Pt/SnO_2 sensing film, two with Au/SnO_2 sensing film and the other two with Pt/Cu(0.16 wt%)-SnO_2. The sensors' operating temperature was 300^oC for the purpose of good sen-

sitivity to studied gases. The sensors output are raw voltage measurements in the form of exponential-like curves, as shown in Figure 2. Data is collected at sampling period of 3sec.

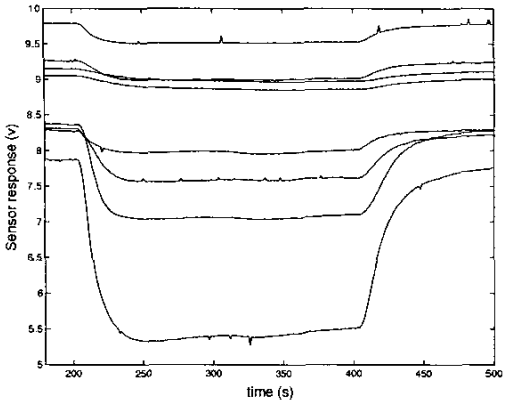

Fig. 2. Raw response of an array of 8 microelectronic gas sensors.

Gases used in the experiment are methane, carbon monoxide, hydrogen, and two binary mixtures: one of methane and carbon monoxide and another of hydrogen and carbon monoxide. Vapours were injected into the gas chamber at a flow rate determined by the mass flow controllers (MFC). Concentration ranges are reported in Table I.

Gas	Concentration range (ppm)
CO	25-200
CH_4	500-4000
CO & CH_4	25-200 & 500-4000
H_2	500-2000
CO & H_2	25-200 & 500-2000

TABLE I
GASES AND THEIR CONCENTRATION RANGES.

The steady state value was recorded for each concentration of the five gases. Figure 2 shows the typical steady state response for the sensor array exposed to different gases. Since our goal is the qualitative classification of patterns, a normalisation procedure is used in order to reduce the influence of concentrations and nonlinearities. Each input pattern is divided by its Euclidean norm. A gas dataset of 220 patterns was used for estimating the performance of different classifiers.

III. FEATURE EXTRACTION

The recognition procedure involves preprocessing and dimensionality reduction, learning from data and validation stage. In order to avoid problems associated with high dimensionality and redundancy, the initial feature vector is projected onto a lower dimensional space. The goal of feature extraction is to find a low dimensional vector that preserves most of

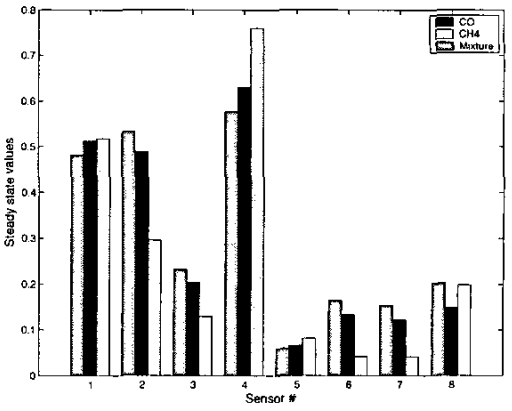

Fig. 3. Histograms showing the response patterns of the eight gas sensors exposed to CH_4, CO and their mixture.

the information in the original feature vector. Most feature extraction techniques for electronic nose applications have been based on linear techniques mainly Principal Components Analysis (PCA).

We used PCA, and compared it with Linear Discriminant Analysis (LDA) and Neuroscale [5]. PCA, a classical linear technique, serves for redundancy removing and feature reduction before the use of the classifier. Figure 4 presents the two-dimensional PCA scores for all the studied gas sensors steady state voltage. We can note that the decision boundaries are not well defined due to strong overlappings.

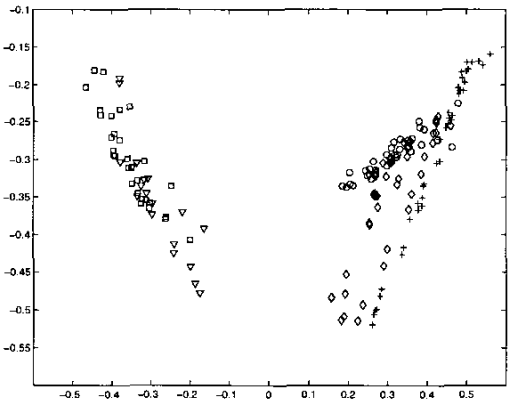

Fig. 4. PCA results for the microelectronic sensor array steady state voltage. Measurement type, CO (circles), CH_4 (plus signs), mixture CO-CH_4 (diamonds), H_2 (triangles) and mixture CO-H_2 (squares).

As compared to PCA, LDA provides a linear projection of the data with (number of classes-1) dimensions, by taking account of the scatter of data within each class and across classes. Figure 5 shows the LDA plots for the studied gases. Compared to PCA results, it is clear from Figure 5, that the LDA method permits to reduce the overlappings between the classes.

Neuroscale projection is a non-linear topographic (i.e. distance preserving) projection that uses a RBF network. This method

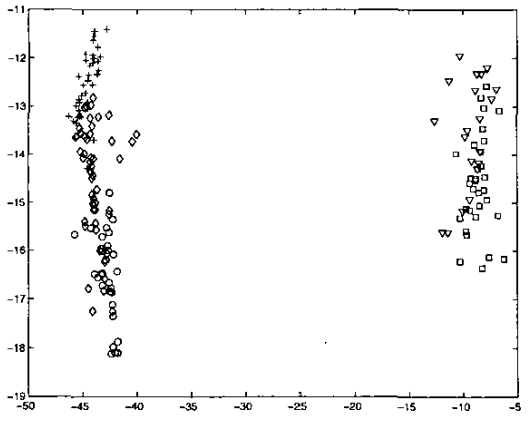

Fig. 5. LDA results for the microelectronic sensor array steady state voltage. Measurement type, CO (circles), CH_4 (plus signs), mixture CO-CH_4 (diamonds), H_2 (triangles) and mixture CO-H_2 (squares).

presents the advantage of preserving the data structure, as well as the possibility of incorporating subjective information. In our case, Neuroscale was used with class information in order to generate a useful feature space that separate classes. Figure 6 shows the Neuroscale plots for the studied gases. Compared to PCA and LDA results, Neuroscale presents the most complex scatter plot with large intra-class variability compared to a small inter-class separation. It is important to notice that the dimension of the projected data is a variable to be optimized.

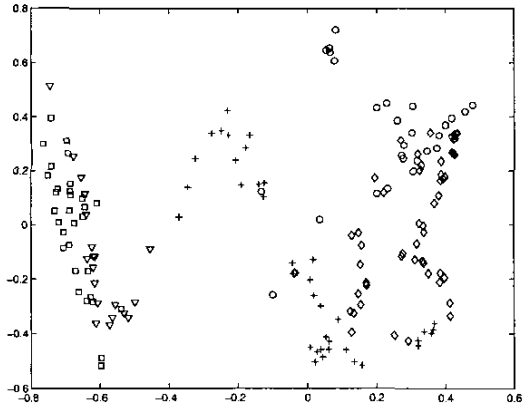

Fig. 6. Data projected using Neuroscale with class information. Measurement type, CO (circles), CH_4 (plus signs), mixture CO-CH_4 (diamonds), H_2 (triangles) and mixture CO-H_2 (squares).

IV. CLASSIFIERS DESCRIPTION

The objective of pattern recognition is to set a decision rule, which optimally partitions the data space into c regions, one for each class C_k. A pattern classifier generates a class label for an unknown feature vector $\boldsymbol{x} \in R^d$ from a discrete set of previously learned classes. The most general classification approach is to use the posterior probability of class membership $\wp(C_k|\boldsymbol{x})$. To minimize the probability of misclassification one should consider the maximum a posterior rule and assign $\boldsymbol{x}$ to class $C_{\hat{k}}$:

$$C_{\hat{k}} = \arg\max_k[\wp(C_k|\boldsymbol{x})] = \arg\max_k[\wp(\boldsymbol{x}|C_k)\wp(C_k)] \quad (1)$$

where $\wp(\boldsymbol{x}|C_k)$ is the class-conditional density and $\wp(C_k)$ is the prior probability. In the absence of prior knowledge, $\wp(C_k)$ can be approximated by the relative frequency of examples in the dataset. One way to build a classifier is to estimate the class-conditional densities by using representation models for how each pattern class populates the feature space. In this approach, classifier systems are built by considering each of the classes in turn, and estimating the corresponding class-conditional densities $\wp(\boldsymbol{x}|C_k)$ from data.

The most widely used method of nonparametric density estimation is the K Nearset Neighbors (KNN). Despite the simplicity of the algorithm, it often performs very well and is an important benchmark method. However, one drawback of KNN is that all the training data must be stored, and a large amount of processing is needed to evaluate the density for a new input pattern. An alternative is to combine the advantages of both parametric and nonparametric methods, by allowing a very general class of functional forms in which the number of adaptive parameters can be increased to build more flexible models. This leads us to a powerful technique for density estimation, called mixture model [6]. In our work we focus on semiparametric models based on mixture distributions. In a mixture model, a probability density function is expressed as a linear combination of basis functions. A model with M components is described as mixture distribution [6]:

$$\wp(\boldsymbol{x}) = \sum_{j=1}^{M} \wp(j)\wp(\boldsymbol{x}|j) \quad (2)$$

where $\wp(j)$ are the mixing coefficients and the parameters of the component density functions $\wp(\boldsymbol{x}|j)$ vary with j. Each mixture component is defined by a Gaussian parametric distribution in d dimensional space. The method for training mixture model is based on maximizing the data likelihood. The log likelihood of the dataset $(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, which is treated as an error, is defined by:

$$l = \sum_{i=1}^{n} \log \wp(\boldsymbol{x}_i) \quad (3)$$

A specialized method is commonly used to produce optimum parameters, known as the expectation-maximization (EM) algorithm [8].

The density models used were Gaussian mixture models (GMM), Generative Topographic Mapping (GTM) [7] and KNN. The discriminant functions used were Generalized Linear Model (GLM), MLP and RBF neural networks [8].

V. CLASSIFICATION RESULTS

Let us now study the classification capabilities of each selected classifier and compare their performances. Experiments are based on our gas sensors data of 220 patterns. Each pattern consists of 8 steady state responses. Since the dataset we used was small, generalization performances were estimated by using the 10-fold cross validation approach. The inputs to each classifier are the projections of the data using PCA, LDA or Neuroscale. The parameters of each mixture models were adapted to the training data in the maximum likelihood framework using EM algorithm. All model selection (for example, the number of hidden units) was chosen according to the validation stage. Tables II and III report the classification performance of the trained classifiers. These tables show that the best results are given by the PCA projection. The most accurate discriminant function is the MLP, while the most accurate density model is the GMM.

Classifier	PCA (%)	Neuroscale (%)	LDA (%)
KNN	89	86.8	83.2
GMM	**92.7**	**88.6**	**90.5**
GTM	89.5	85.5	83.2

TABLE II
CLASSIFICATION PERFORMANCE FOR DENSITY MODELS, USING PCA, NEUROSCALE AND LDA PROJECTIONS.

Classifier	PCA (%)	Neuroscale (%)	LDA (%)
RBF	83	76.8	80
MLP	**90**	**88.2**	**89.4**
GLM	86.4	86.4	81.8

TABLE III
CLASSIFICATION PERFORMANCE FOR DISCRIMINANT FUNCTIONS, USING PCA, NEUROSCALE AND LDA PROJECTIONS.

For the PCA tool, the best performance is achieved when projecting to five principal components. The addition of components actually degrades the performance of the classifier as shown in Figure 7. GMM is shown to outperform MLP.

VI. CONCLUDING REMARKS

In this paper we presented a gas identification method based on either class-conditional density estimation or discriminant classifier. We conducted comparative experiments to decide which algorithms were most suitable for the problem of classifying combustion gases. Data were obtained from an array of 8 Microelectronic gas sensors using an experimental setup. It was found that GMM offers the best performance as compared to MLP, GTM, KNN, RBF and GLM. A significant dependence of classifiers performance on the number of principal components was found.

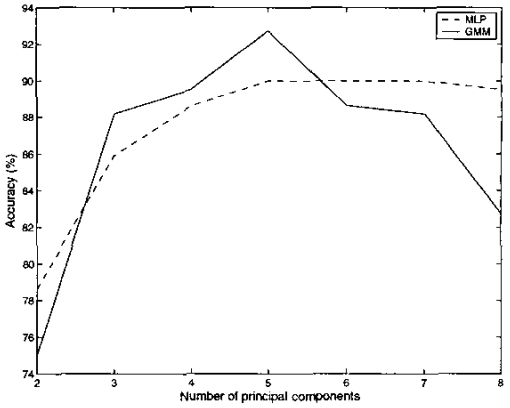

Fig. 7. Accuracy as a function of the number of principal components.

Work is underway in order to incorporate more features than the steady state value. Pattern analysis systems will be designed to classify and quantify the different gases.

ACKNOWLEDGMENTS

The work described in this paper was supported by a Direct Allocation Grant (project No. DAG02/03.EG05) and a PDF Grant for HKUST.

REFERENCES

[1] P. C.H. Chan, G. Yan, L. Sheng, R. K. Sharma, Z. Tang, J.K.O. Sin, I-M. Hsing, Y. Wang, "An integrated gas sensor technology using surface micro-machining," Sensors and Actuators B 82, pp. 277-283, 2002.
[2] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: a review," IEEE Sensors Journal, Vol. 2, NO. 3, pp. 189-202, 2002.
[3] Y. Zhang, M. Alder and R. Togneri, "Using Gaussian mixture modeling in speech recognition," In Proc. ICASSP Conf., 1994, Vol. 1, pp. 613-616.
[4] N. Vasconcelos and A. Lippman, "Feature representations for image retrieval: beyond the color histogram," In Proc. IEEE ICME Conf., 2000, Vol. 2, pp. 899-902.
[5] D. Lowe and M. E. Tipping, "Feed-forward neural networks and topographic mappings for exploratory data analysis," Neural Computing and Applications, 4, pp. 83-95, 1996.
[6] D.M. Titterington, A.F.M. Smith and U.E. Makov, "Statistical analysis of finite mixture distributions," John Wiley, New York, 1985.
[7] C. M. Bishop, M. Svensen, and C.K.I. William, "GTM: The Generative topographic mapping," Neural Computation, 10 (1), pp. 215-235, 1996.
[8] C. M. Bishop, "Neural networks for pattern recognition," Clarendon press, Oxford, 1995.