GAS IDENTIFICATION WITH MICROELECTRONIC GAS SENSOR IN PRESENCE OF DRIFT USING ROBUST GMM

Sofiane Brahim-Belhouari, Amine Bermak and Philip C. H. Chan

Hong Kong University of Science and Technology,
Electrical and Electronic Engineering Department
Clear Water Bay, Kowloon, Hong Kong, SAR.

ABSTRACT

The pattern recognition problem for real life applications of gas identification is particularly challenging due to the small amount of data available and the temporal variability of the instrument mainly caused by drift. In this paper we present a gas identification approach based on class-conditional density estimation using Gaussian mixture models (GMM). A drift counteraction approach based on extracting robust feature using a simulated drift is proposed. The performance of the retrained GMM shows the effectiveness of the new approach in improving the classification performance in the presence of artificial drift.

1. INTRODUCTION

Gas identification on a real-time basis is very critical for a very wide range of applications in the civil and military environments. The past decade has seen a significant increase in the application of multi-sensor arrays to gas classification and quantification. Most of this work has been focused on systems using microelectronic gas sensors featuring small size and low-cost fabrication, making them attractive for consumer applications. A number of interesting applications have also emerged in the last decade whether related to hazard detection, poisonous and dangerous gases or to quality and environmental applications such as air quality control. Furthermore, in the next decade, increased interest in mass production of reliable microelectronic gas sensors is expected in automation process manufacturing as well as in the automotive industry as recent requirements for active protection of passengers in vehicles from inherent pollutants are being introduced. Unfortunately, the present gas sensors have a lack of selectivity and therefore respond similarly to a wide variety of gases. Figure 1 illustrates an example of a microelectronic gas sensor's response to carbon monoxide CO, with 4 different concentrations of methane CH_4. It can be shown from this simple illustration that different CO concentrations may result in similar voltages at the output of the sensor depending on the concentration of the interfering gas. This makes the problem of detecting and further quantifying a gas mixture a challenging task.

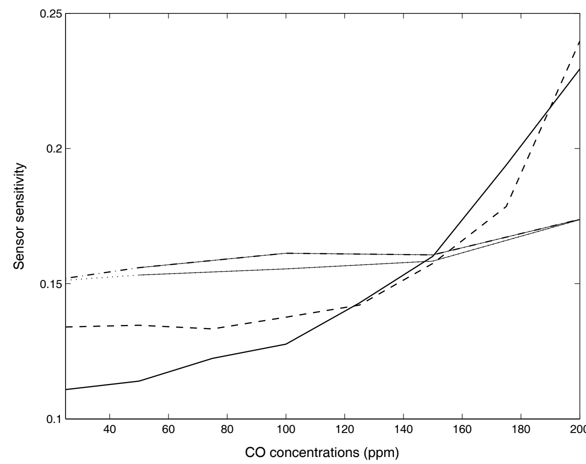

Fig. 1. Sensor responses as a function of CO concentration for different concentrations of CH_4.

The idea to combine an array of sensors with a pattern recognition algorithm to improve the selectivity of the single gas sensor has been widely accepted and being used by researchers in this field. In fact, an array of different gas sensors is used to generate a unique signature for each gas. 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 research work has been carried out during the last decade in gas detection, preprocessing algorithms as well as pattern recognition classification [1] [2]. Unfortunately, the relative success of gas identification systems using the pattern recognition algorithms is primarily based upon laboratory measurements in well-controlled environments. Among the most serious limitation of actual gas classification and quantification systems is the inherent drift of gas sensors, which shows significant temporal variations of the sensor response when exposed to identical atmospheres. These drifts are due to unknown dynamic processes in the sensor system (e.g. poisoning or ageing of sensors) or environmental changes (e.g. temperature and pressure conditions). As a result of drift, the cluster distribution in the feature space becomes unstable over time, making useless the decision surface built by the classifier during the training phase. Therefore, after certain time, drift impairs the pattern recognition ability of the system, which needs to be completely retrained.

In this paper we will present a gas classification approach based on class-conditional density estimation using Gaussian mixture models (GMM). The proposed classifier is found to outperform both Multi Layer Perceptron (MLP) and K Nearest Neighbors (KNN). A drift counteraction approach based on extending the training data set using a simulated drift is proposed. The performance of the retrained GMM shows the effectiveness of the new approach in improving the classification performance in the presence of additive as well as multiplicative drift. Section 2 of the paper briefly presents the theoretical background of GMM classifier. Section 3 describes the classification results with and without drift using GMM classifier. This section also describes the experimental set-up used to sample the data from the microelectronic gas sensor array. Section 4 concludes the paper.

2. GAUSSIAN MIXTURE MODELS

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 class 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 Nearest 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 [3]. In our work we focus on semiparametric models based on Gaussian mixture distributions.
In a Gaussian mixture model, a probability density function is expressed as a linear combination of basis functions. A model with M components is decribed as mixture distribution [3]:

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

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:

$$\wp(\boldsymbol{x}|j) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_j|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\}$$

The parameters to be estimated are the mixing coefficients $\wp(j)$, the covariance matrix $\boldsymbol{\Sigma}_j$ and the mean vector $\boldsymbol{\mu}_j$. 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-maximisation (EM) algorithm [4].

3. EXPERIMENTAL RESULTS

3.1. 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 2). The sensor array composed of 8

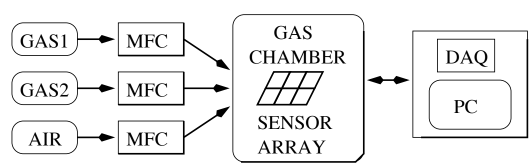

Fig. 2. Scheme of the experimental setup. MFC stands for Mass Flow Controller.

micro-hotplate based SnO_2 thin film gas sensors, have been used [5]. 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 chosen to be 300^oC for the purpose of good sensitivity to the studied gases. The sensors output are raw voltage measurements in the form of exponential-like curves, as shown in Figure 3. Gases used in the experiment are methane, car-

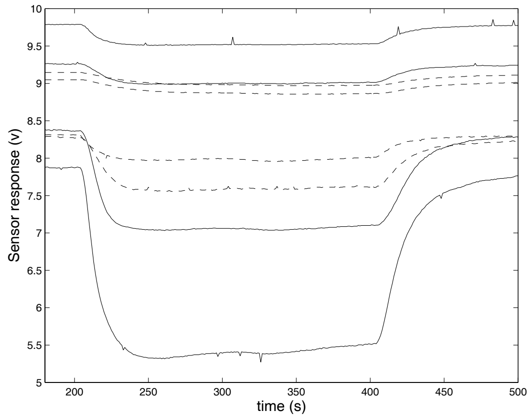

Fig. 3. Raw response of an array of 8 microelectronic gas sensors.

bon monoxide, hydrogen, and two binary mixtures: one of methane and carbon monoxide and another of hydrogen and carbon monoxide. Vapors were injected into the gas chamber at a flow rate determined by the mass flow controllers (MFC). Concentration ranges are reported in Table 1. The

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 1. Gases and their concentration ranges.

steady state values of the array sensor were recorded while periodically injecting different gases. A gas data set of 220 examples was created to evaluate the performance of different pattern recognition systems. Each example consists of 8 sensor transients, with $i = 1, ..., N_T$ samples per transient denoted by $v(t_i)$. We subtracted the baseline of each sensor in order to reduce the effects of the additive sensor drift. Since our goal is the qualitative classification of patterns, a normalization procedure is used in order to reduce the influence of concentrations and non-linearities. Each input pattern is divided by its Euclidean norm.

3.2. Classification without drift

Prior to applying the GMM classifier, a dimensionality reduction technique namely PCA was used in order to perform redundancy removing and feature reduction. 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 a strong overlapping.

In order to compare the performance of different classifiers, a 10-fold cross validation approach was used allowing to overcome the problem of the limited data set typically available in gas sensors applications. It is well known that

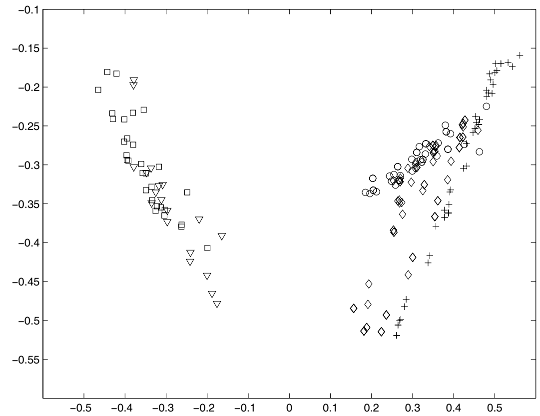

Fig. 4. PCA results for the microelectronic sensor array steady state voltage. Measurement type, CO (circles), CH_4 (plus signs), mixture CO-CH_4 (diamonds), H_2 (triangles) and mixture CO-H_2 (squares).

the performance of a given classifier depends on the number of principal components. In order to obtain a more objective comparison, we reported in Figure 5, the classification success of GMM, KNN and MLP as function of the number of principal components. The best performance is achieved using GMM with a success rate of 92.7% obtained for 5 principal components.

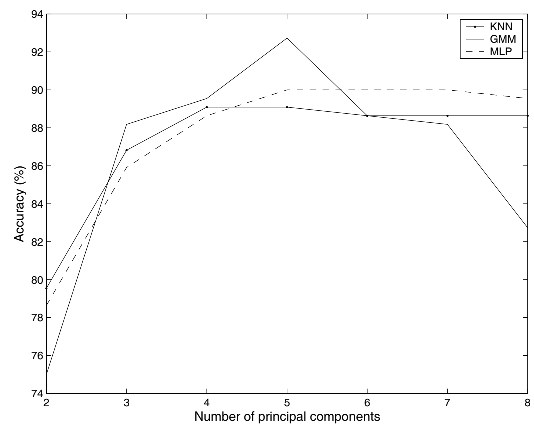

Fig. 5. Accuracy as a function of the number of principal components.

3.3. Classification with drift

Among the most serious limitation of actual gas sensors is the drift problem, which shows significant temporal variations of the sensor response when exposed to identical atmospheres. Drift problem can be explained as a random temporal variation of the sensor response when exposed to the same gases under identical conditions. It can affect both the baseline (additive) and the sensitivity of the sensor (multiplicative). Figure 6 illustrates an example of an additive drift problem in which we have reported the real response of the sensor as function of the concentration of gases periodically injected into a gas chamber in which the sensors are being placed. We can note that the baseline response of the sensor is shifted which complicates the classification problem even further.

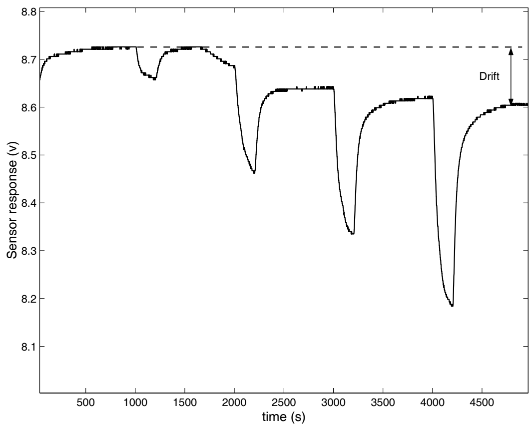

Fig. 6. Additive drift affecting the sensor baseline.

The drift causes temporal variations of the pattern distribution in the feature space. This makes obsolete the decision surface obtained during the training phase and hence retraining the entire system is necessary.

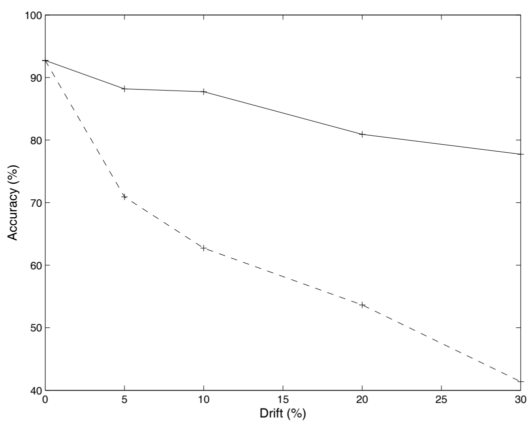

Fig. 7. Classification performance as function of drift (expressed in %) before (dashed) and after (solid) retraining.

To compensate for the patterns dispersion movement, we propose to extract robust features by generating simulated drift. The efficiency of this procedure has been tested against simulated linear drift. The drift has been modelled as $v_d(t) = v(1+\alpha t)$ where v is the sensor output before the drift experiment and αt_{max} was chosen randomly for each sensor [6]. Drift varying between 0 and 30% has been artificially generated. The performance of the best classifier was evaluated over the drifted measurements. Figure 7 shows that drift affects the recognition ability of the GMM as the classification success declines significantly (dashed line of figure 7). The drift counteraction strategy is to retrain GMM using drifted sensor responses (solid line of figure 7). The performance of the retrained GMM was evaluated using the 10-fold cross validation method. It is shown that the counteraction procedure improves the performance of GMM in presence of 30% drift by a factor of over 35%. The final assessment of this procedure has to be achieved by testing it over real sensor's drift data.

4. CONCLUSION

In this paper we presented a gas identification approach based on class-conditional density estimation using Gaussian mixture models (GMM). The proposed classifier is shown to outperform both KNN and MLP for gas sensors data set collected from an integrated gas sensor array. It was however found that the drift seriously degrades the classification performance of GMM. A drift counteraction approach based on extracting robust feature using a simulated drift was proposed. The performance of the retrained GMM was evaluated using a cross validation method which shows a gain of over 35% obtained for up to 30% drift. The test of the proposed approach for real drift values obtained for example by varying the operating temperatures is underway.

Acknowledgments

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

5. REFERENCES

[1] R. Gutierrez-Osuna, “Pattern analysis for machine olfaction: a review,” IEEE Sensors Journal, Vol. 2, NO. 3, pp. 189-202, 2002.

[2] M. Pardo and G. Sberveglieri, ”Learning from data: A tutorial with emphasis on modern pattern recognition methods”, IEEE Sensors Journal, vol. 2, NO. 3, pp. 189-202, 2002.

[3] D.M. Titterington, A.F.M. Smith and U.E. Makov, “Statistical analysis of finite mixture distributions,” John Wiley, New York, 1985.

[4] C. M. Bishop, “Neural networks for pattern recognition,” Clarendon press, Oxford, 1995.

[5] 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 micromachining,” Sensors and Actuators B 82, pp. 277-283, 2002.

[6] Marco, S.; Ortega, A.; Pardo, A.; Samitier, J., “Gas identification with tin oxide sensor array and self-organizing maps: adaptive correction of sensor drifts,” IEEE Transactions on Instrumentation and Measurement, Volume: 47 Issue: 1 , PP. 316 -321, Feb. 1998.