

# New approach for gas identification using supervised learning methods (SVM and LVQ)

*Aref Harakeh*<sup>1,2</sup>, *Samia Mellah*<sup>2</sup>, *Mustapha Ouladsine*<sup>2</sup>, *Rafic Younes*<sup>3</sup> and *Catherine Bellet*<sup>1</sup>

<sup>1</sup>PARTELEC, 29 Rue Saint Lazare, 77170 Brie Compte Robert, France

<sup>2</sup>LIS UMR CNRS 7020, 52 Av. Escadrille Normandie Niemen, 13397 Marseille Cedex 20, France

<sup>3</sup>Université Libanaise, Faculté du Génie, Hadat, Baabda, Liban

**Abstract.** This article proposes a new approach for gas identification, this approach relies on applying supervised learning methods to identify a single gas as well as a mixture of two gases. The gas is trapped in a gas discharge tube, it is then ionized at a relatively low pressure using an HV transformer. The images captured after the ionization of each single gas is then captured and transformed into a database after being treated in order to be classified. The obtained results were very satisfying for SVM as well as for LVQ. For the case of identification of a single gas, the learning rate as well as the validation rate for both methods were 100%. However, for the case of mixture of two gases, a Multi-Layer Perceptron neural network was used to identify the gases, the learning rate as well as the validation rate were 98.59% and 98.77% respectively. The program developed on MATLAB takes the captured image as an input and outputs the identified gases for the user. The gases used in the experiments are Argon (Ar), oxygen (O<sub>2</sub>), Helium (He) and carbon dioxide (CO<sub>2</sub>).

## 1 Introduction

Gas sensors and detectors are essential devices in many industrial applications. The aim of this article is to present a new approach that uses supervised learning methods in order to detect a mono gas as well as to recognize the presence of two gases in a mixture. This approach could be considered as an important step toward developing another new approach for identifying several gases in a mixture (more than two).

## 2 State of the art:

Machine Learning is an application of artificial intelligence that affords to systems the ability to learn from a given data (experience) without being explicitly programmed. In this part of the article, we explain briefly the theory of the supervised learning methods (SVM and LVQ) that were used for gas identification.

**2.1 SVM (Support Vector Machines):** A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given

labelled training data (supervised learning), the role of the algorithm is to output an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side [1].

**2.2 LVQ (Learning Vector Quantization):** LVQ algorithms are related to other competitive learning algorithms such as self-organizing maps (SOMs) [2] and c-means. Competitive learning algorithms are based on the winner-take-all learning rule, and variants in which only certain elements or neighbourhoods are updated during learning. The original LVQ algorithms and most modern extensions use supervised learning for obtaining class-labelled prototypes (classifiers). However, LVQ can also be trained without labels by unsupervised learning for clustering purposes. [3]

### 3 Results and Discussion:

The database obtained from the response of the mono gas (images captured after ionization of each gas) is of dimension (Mx4). The first three columns correspond to the three variables RGB characterising each sample  $x_i$  (pixel). The fourth corresponds to the class of the sample  $y_i$ . M is the number of samples. Figure one shows the images captured from ionization of the four gases Argon (Ar), CO<sub>2</sub>, Helium (He) and O<sub>2</sub>.

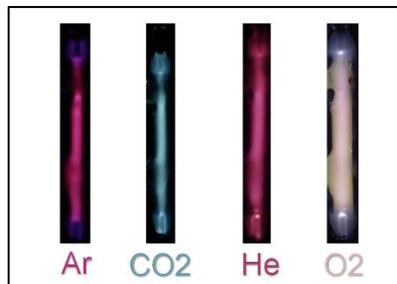


Figure 1: Spectrum of Ar, CO<sub>2</sub>, He and O<sub>2</sub>

The data are three-dimensional, hence tracing the points  $x_i$  in a 3D space gives us an idea about the linear or non-linear separability of the data. The repartition of this data in 3D space shows us that the data is totally linearly separable, hence the usage of artificial intelligence or SVM approaches are well adequate to this problem.

#### 3.1 Identification by SVM:

Since the problem is a multi-class problem (4 classes), we used the tools of SVM multi-classes. We chose the one against one algorithm. We are searching for the number of hyperplanes. Using each time as input the samples of the two classes to separate. To classify a new spectrum (identify the gas corresponding to it), a majority vote on the six classifiers is implemented. As we saw when viewing the database, the data is perfectly linearly separable. Therefore, the application of hard-margin SVMs is sufficient to find the optimal hyperplanes that perfectly divide this data into four classes. Then the matrix of quadratic coefficients Q is generated. The criteria to minimise is written in the quadratic program as:

$$\min \frac{1}{2} \alpha^T Q \alpha + f^T \alpha \quad (1)$$

Respecting the constraints:  $y^T \alpha = 0$  (2) and  $0 \leq \alpha_n \leq C$  (3)

With:  $f = -(\text{unity vector of dimension } (N \times 1))$  and  $C = \infty$

The command “quadprog” in Matlab permits us to solve the minimisation problem with constraints. As mentioned earlier, we use the one against one approach, so for every two different data classes we follow these steps in order to determine the hyperplane that separates their data. Thus, we will get six hyperplanes. To classify our data with the SVM approach, we first tested the SVM hard margin method. The result obtained is very satisfying (learning and validation rate 100%).

### 3.2 Identification by LVQ:

The architecture of the network can vary according to the number of neurons of the competitive layer, and according to the number of neurons of this same layer assigned to each class. The purpose of learning is the estimation of the coefficients of the matrix  $W$  (connection weights) to best fulfil the task for which the network is intended. By phase of learning is therefore meant the phase in which the network coefficients undergo modifications according to rules called learning rules, until the network is stabilized, that is to say until what the desired output is almost achieved. The parameters to be optimized are generally four in number: the number of neurons of the competitive layer and the number of neurons affected at each class, the number of iterations, adjustment coefficient, and the coefficient of initialisation of weight. The resultant configuration for this network gives us a learning rate and a validation rate equals to 100%.

### 3.3 Identification of Two Gases using an MLP Neural Network:

In our case (four gases), there exist six possible cases for mixtures of two gases which are (CO<sub>2</sub>,O<sub>2</sub>), (CO<sub>2</sub>,Ar), (CO<sub>2</sub>,He), (O<sub>2</sub>,Ar), (O<sub>2</sub>,He), and (Ar,He). For every case of mixture of two gases, the total RGB of all possible mixtures ranging from (99% Gas1, 1% Gas2) to (1% Gas1, 99% Gas2) was provided as an input database to the Matlab neural network toolbox (nftool). The output corresponding to the total RGB for mixtures of (CO<sub>2</sub>,O<sub>2</sub>), (CO<sub>2</sub>,Ar), (CO<sub>2</sub>,He), (O<sub>2</sub>,Ar), (O<sub>2</sub>,He), (Ar,He) are 1100, 1010,1001,0110,0101,0011 respectively. The neural network was trained on 70% of the input samples. Validation and testing were done on 15% of input samples. The learning rate as well as the validation rate were 98.59% and 98.77% respectively. The figure below shows a Matlab GUI that takes the measured values of RGB as an input, and uses the neural network to identify the gases in the mixture.



Figure 2: Matlab GUI for gas identification

## 4. Conclusion and Perspectives

As a result, a new approach for gas identification using SVM and LVQ was studied and tested on four different gases. This approach allows us to identify only a mono gas, the results showed a 100% learning rate for SVM and LVQ, which is considered to be promising. In addition, an MLP neural network was used in order to identify the gases in a mixture of two

gases, the learning rate as well as the validation rate were 98.59% and 98.77% respectively. It will be more interesting in future work to develop an advanced algorithm capable to identify several gases (more than two) in a mixture. This requires a much bigger database that provides more information for the machine allowing it to identify gases in a mixture. The proposed approach is essential to go further in identifying gases in a mixture.

## References

- [1] Marsland, S. (2011). *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC.
- [2] Kohonen T (1997) Self-organizing maps. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [3] Bezdek JC, Pal NR (1995) Two soft relatives of learning vector quantization. *Neural networks* 8(5):729–743