Machine Learning Methods in Electronic Nose Analysis

Irene Rodriguez-Lujan
Universidad Autónoma de Madrid,
irene.rodriguez@uam.es

Jordi Fonollosa
Institute for Bioengineering of Catalunya,
jfonollosa@ibecbarcelona.eu

Ramon Huerta
University of California San Diego,
rhuerta@ucsd.edu

***Abstract*—The main existent tool to monitor chemical environments in a continuous mode is gas sensor arrays, which have been popularized as electronic noses (enoses). To design and validate these monitoring systems, it is necessary to make use of machine learning techniques to deal with large amounts of heterogeneous data and extract useful information from them. Therefore, enose data present several challenges for each of the steps involved in the design of a machine learning system. Some of the machine learning tasks involved in this area of research include generation of operational patterns, detection anomalies, or classification and discrimination of events. In this work, we will review some of the machine learning approaches adopted in the literature for enose data analysis, and their application to three different tasks: single gas classification under tightly-controlled operating conditions, gas binary mixtures classification in a wind tunnel with two independent gas sources, and human activity monitoring in a NASA spacecraft cabin simulator.**

I. Introduction

Gas sensors can detect changes in temperature, humidity, air pressure, and, obviously, human presence. Ogawa and Togawa [1] show that they are able to identify specific events in a home like waking up, start to cook, have breakfast, go out, come home, have supper, and go to bed. These observations hinted at the use of chemical traces as a manner to monitor homes [2], but the possibilities of environment monitoring are immense since these sensors can be used to monitor many other places such as hospitals or offices. In any of these applications, the use of gas sensor arrays requires gas sensor calibration. The calibration of a gas sensor array consists on establishing the functional relationship between measured values, analyte quantities and/or analyte identification. Traditionally, calibration includes first, the selection of the functional form of a computational model; second, the estimation of the corresponding model parameters based on a training dataset; and third, the model validation [3]. The resulting computational model is then used to predict the analyte amount/class of new measurements. Therefore, gas sensor calibration naturally translates to classification and regression problems in machine learning, and the sequence of steps involved in the design of a machine learning system must be taken, namely: data collection, features and model (classification/regression) selection, and training and evaluation [4]. In this work, we will focus on machine learning methods used for enose calibration in four different applications (I.1, I.2, II.1, III.1) that make use of three sources of data (I-III):

I. Electronic nose data under tightly-controlled operating conditions [5]. An array of 16 metal-oxide (MOX) gas sensors is exposed to six different volatile organic compounds at different concentration levels under tightly-controlled operating conditions. An extensive dataset was collected over a period of three years ($13,910$ measurements). For each measurement, a 128-component vector is processed from the sensors' responses to extract steady-state and transient features (Section III).
   1. Active sensor calibration for the discrimination of three gases [6]. This work investigates the optimal experiment selection to calibrate a gas sensor array to get the maximal possible performance in the classification of three different gases. It will be described in more detail in Section II.
   2. Sensor calibration for the discrimination of six gases [7]. This work proposes a new pointwise Fisher consistent multiclass hinge loss function, which is used to efficiently calibrate an electronic nose and classify six different gases.

II. Electronic nose data for turbulent gas mixtures [8]. In order to reproduce more realistic environments, an electronic nose composed of 8 chemo-resistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel with two independent gas sources that generate two gas plumes. The sensor array was exposed to binary mixtures of ethylene with either methane or carbon monoxide. Volatiles were released at four different rates to induce different concentration levels in the vicinity of the sensor array. Each measurement is defined by 8 time series corresponding to MOX sensors' conductivity together with temperature and humidity information.
   1. Sensor calibration for ethylene discrimination in binary gas mixtures [9]. This work proposes a sensor calibration methodology to detect ethylene in a turbulent and changing background composed of methane or carbon monoxide on air (binary classification).

III. Electronic nose data from a NASA spacecraft cabin simulator. An electronic nose was installed and operated continuously while different volunteers (15 females and 47 males) were performed different daily activities (physical exercise on fitness equipment, defrosting frozen dinners, eating dinner, and hygienic activities). This dataset has

sensor readings for four weeks: two weeks used for training and two weeks used for testing. Data are captured at 0.05Hz and each pattern represents one-hour time series. Therefore, the dataset is formed by 672 patterns. Each measurement is defined in a $4,320$ dimensional space associated with 24 180-dimensional time series corresponding to 24 polymeric sensors' responses in one hour.

1. Sensor calibration for human activity monitoring [10]. This paper performs enose calibration to predict the number of people in the room (regression).

In the following sections, we will revisit some of the machine learning approaches adopted in these applications and involving different steps of a machine learning system. In section II, the data collection procedure is analyzed to demonstrate the impact that external control parameters (gas concentrations) can have on the performance of a calibrated system. Section III will describe different techniques for feature selection and extraction used to build the attribute space that will serve as input to the classification/regression model. Section IV will focus on presenting some of the classification and regression models used for sensor calibration. Finally, Section V will point out some further work directions.

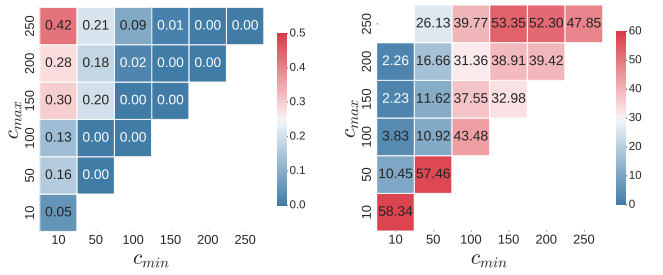
II. Data Collection: Optimal Experiment Selection

A training dataset needs to be collected to perform the calibration of an analytical system. The challenge facing engineers building sensor arrays for event discrimination is that data is not available until the particular experiment is conducted. Moreover, each experiment carries a significant overhead that has no guarantee of positively impacting calibration quality. For example, each of the $13,910$ experiments in the chemical gas sensor array dataset described in [5] requires 20 minutes of sensor array exposure time to collect just one example and, depending on the experiment, it may also carry health risks to those in the lab. Additionally, systems based on metal-oxide gas sensors are dynamic systems that need to be periodically recalibrated [11]–[13]. These difficulties are present in many applications including space travel [10], [12], environmental monitoring of public spaces [1], and industrial leak detection [14], making it fertile ground for the machine learning community.

In order to alleviate the experimental cost and reduce the frequency of the recalibrations, a methodology to select the best training examples to calibrate the system without modifying the configuration of the sensor array is proposed in [6]. This approach differs from other approaches in the literature focused on modulating sensors' parameters such as frequencies, operating voltages, or temperature in response to environmental changes or application needs [15], [16]. In [6], the focus is on sensor networks where one has has to calibrate a device in controlled conditions and then deploy it for real operation. Given an analyte discrimination problem in which every training sample (measurement) is defined by a class label (gas type) and a control parameter c (gas concentration), the authors analyze which sampling distribution must follow the control parameter in the next batch of experiments in order to obtain the best calibration of the sensor array at each time, so the classifier is trained sooner rather than later. Note that the concentration level of each recording is not explicitly provided to the classifier. The active sampling strategy focuses on gas concentration selection since it is known that classification performance is more heavily impacted by concentration selection than gas label selection [17], [18]. They assume a canonical sampling probability distribution over the space of gas concentrations $p^{\kappa}(c) \propto \exp(-\kappa c)$ with κ the rate of the sampling distribution. This distribution allows biasing the sampling towards lower ($\kappa > 0$) to higher ($\kappa < 0$) concentrations, and it also recovers the uniform (uninformed) distribution ($\kappa = 0$). The proposed active sampling strategy works at batch level, and a fixed value for κ is set for each batch of size B. Then, given the best sampling strategy in the preceding batches and given the sampling distribution $p^{\kappa}(c)$, the algorithm selects the optimal κ value for the following batch experiments as that minimizing the cross-validation error of an Inhibitory Support Vector Machine classifier (ISVM) [19] defined by the cost parameter C and the inverse of the kernel width γ. This classifier is described in more detail in Section IV. Under this framework, the probability of sampling at concentration c at the m-th batch given the sequence of the optimal κ_i up to the m-th batch, $\{\kappa_1, \kappa_2, \ldots, \kappa_m\}$, is given by the following equation

$$p_m\left(c|\kappa_1, \kappa_2, \ldots, \kappa_m\right) = \frac{1}{m}\sum_{i=1}^{m} \frac{\exp(-\kappa_i c)}{\sum_{c' \in \mathcal{C}} \exp(-\kappa_i c')}\,, \qquad (1)$$

where $\mathcal{C}$ is the space of feasible values for the gas concentrations. Each addend corresponds to the probability at level c for one batch by having normalized the sampling distribution $exp(-\kappa c)$ so that $\sum_{c' \in \mathcal{C}} p(c) = 1$. Finally, the factor $\frac{1}{m}$ comes from the normalization of the joint probability $p_m\left(c|\kappa_1, \kappa_2, \ldots, \kappa_m\right)$ so that $\sum_{c' \in \mathcal{C}} p_m\left(c'|\kappa_1, \kappa_2, \ldots, \kappa_m\right) = 1$. The proposed algorithm is shown in Algorithm 1. The algorithm is tested in a subset of the enose Dataset I described in Section I (Application I.1.). The final dataset has $1,800$ 128-dimensional patterns of three distinct pure gaseous substances, namely Ethanol, Ethylene and Acetaldehyde, each dosed at concentration values ranging from 2.5 μmol/mol (ppm) to 300 μmol/mol (ppm). The uniform distribution of the conditional probabilities $P(gas|c)$ in this dataset ensures uninformative (random) sampling on the label. The grid of κ values ranged from -0.03 to 0.03 with a stepwise resolution of 0.001. The results reported in this work show that the active sampling strategy described in Algorithm 1 can only improve an uninformed (random) selection of samples. The best experimental configuration yields a mean classification error of 0.5% as it will be shown in Section IV. Additionally, Fig. 1 shows the classification error in Algorithm 1 with $\kappa = 0$ as a function of the range of concentrations used for training and testing the ISVM model. As expected, the performance of the model is better when



(a) (b)

Fig. 1. Average error rate (%) in the classification of three gases when the electronic nose is calibrated using 200 recordings using Algorithm 1 ($\kappa = 0$) in the range of concentrations $[c_{min}, c_{max}]$ and tested (a) in the same range of concentrations or (b) in a range of concentrations out of range $[c_{min}, c_{max}]$.

the range of concentrations in testing samples is the same as that used for training the classifier. According to Fig. 1(a), the lowest error rates for evaluations in the same range of concentrations are obtained when models are trained in a narrow range of concentrations (around the diagonal). On the contrary, Fig. 1(b) shows that the lowest error rates for evaluations of recordings in a range of concentrations not used in training correspond to scenarios when the sensor is calibrated using the widest range of concentrations. Furthermore, not including low concentrations during calibration significantly worsens the classifier's performance; this is not a surprising result as it is generally believed that gases at higher concentrations are easier to classify than those at lower concentrations.

Algorithm 1 Active sampling algorithm where the main input parameters are the entire dataset U, the rate of the sampling distribution at the current stage κ, the sequence of the optimal $\{\kappa_1, \kappa_2, \ldots, \kappa_m\}$ values for the previous stages, and the batch size B. The output is the average error rate, μ and the standard deviation, σ, for N_{CV} draws using an ISVM as classifier. In all the runs, N_{CV} is set to 100.

Inputs: U, κ, $\{\kappa_1, \kappa_2, \ldots, \kappa_m\}$, B, N_{CV}.
Outputs: μ,σ
for $C \leftarrow [0.1, 1, 10, 100, 1000, 5000, 10000, 50000]$ **do**
  for $\gamma \leftarrow [0.005, 0.01, 0.05, 0.1, 1.0, 5.0, 10.0]$ **do**
    $\mathbf{v} \leftarrow \vec{0}$
    for $t \leftarrow 1 \ldots N_{CV}$ **do**
      $S \leftarrow$ sample $(m+1) \times B$ points according to Eq. 1.
      $T \leftarrow U \setminus S$
      Train the ISVM (C,γ) classifier on S.
      $v_t \leftarrow$ test error on T
    end for
    $\tilde{\mu}_{C,\gamma} \leftarrow$ mean($\mathbf{v}$)
    $\tilde{\sigma}_{C,\gamma} \leftarrow$ std($\mathbf{v}$)
  end for
end for
$\mu \leftarrow \min_{C,\gamma} \tilde{\mu}_{C,\gamma}$
$\sigma \leftarrow \min_{C,\gamma} \tilde{\sigma}_{C,\gamma}$

Finally, it should be also remarked that this idea of optimally selecting the next experiment to calibrate the sensor array is in line with the active sampling algorithms proposed in machine learning [20]. For a more detailed discussion about the proposed algorithm, the experimental results, and the relation between this work and active learning techniques in machine learning , the reader is referred to [6].

III. Feature Extraction and Selection

A critical point for the success of a machine learning system is the correct representation of the input space of the machine learning model. When working with electronic noses based on metal-oxide gas sensors, the sensor response is given by the conductivity across the active layer of each sensor [21]. The interaction processes between the sensor and the analyte identity and/or concentration dosage define the response profile, which allows machine learning algorithms to perform different tasks such as sensor calibration for gas discrimination or activity monitoring. Different strategies can be found in the literature to represent the input space data. Some of them directly use the raw time series provided by the sensors and solve a multidimensional classification/regression problem capable of dealing with the temporal structure of the data [22]. Other approaches downsample the time series to obtain a multivariate representation of the data, and then apply a classification/regression algorithm that does not take into account the temporal structure such as Support Vector Machines with an Gaussian Kernel. This is the approach adopted for sensor calibration in gas binary mixtures (Application II.1) and human monitoring applications (Application III.1) presented in Section I. Finally, new features can be generated based on the sensors' response time series. For example, the sensor calibration problem under tightly-controlled operating conditions introduced in Section I and used in Section II generates eight features for each sensor: two steady-state features and six features reflecting the sensor dynamics. Let $r[k]$ be the sensor resistance time profile, k the discrete time indexing in the recording interval $[0, T]$, and T the duration of the measurement, the steady-state feature ΔR is given by the difference of the maximal resistance change and the baseline, $\Delta R = \max_k r[k] - \min_k r[k]$. Its normalized version $||\Delta R|| = \frac{\Delta R}{\min_k r[k]}$ is also useful for gas discrimination. On the other hand, six features based on the the exponential moving average (ema_α) reflect the sensor dynamics of the increasing/decaying transient portion of the sensor responses [23]. The ema_α transform evaluates the rising/decaying portions of the sensor resistance by considering the maximum/minimum values of a first-order digital filter $y[k] = (1-\alpha)y[k-1] + \alpha\,(r[k] - r[k-1])$, with $y[0] = 0$ and the smoothing parameter $\alpha \in [0, 1]$. Different values of α generate different features with information about the transient response, so the six features are generated by setting $\alpha = \{0.1, 0.01, 0.001\}$ for both the rising and the decaying stages.

Besides the importance of setting a representative input space, another common issue in electronic nose data is that gas sensors are characterized by high correlation in their response [24]. This redundancy can be used to extend the lifetime of

sensor arrays [25], [26] and to correct sensor drift [13], [27]. As an example, Fig. 2 shows the Pearson correlation for the 10×10 pairs of sensors utilizing the complete time series captured for each sensor in Dataset II (Section I). From Fig. 2, we can conclude that the sensors TGS2600 and TGS2602 are highly correlated among them, as well as the sensors TGS2620, TGS2612, TGS2611, and TGS2610. Therefore, if some of these sensors are not considered, the information provided by the rest of the sensor array is expected to be similar to the complete sensor array. That is why, feature selection is a common step in the design of a chemical detection system [28]–[32]. One of the methods used for selecting the optimal subset of sensors is the Quadratic Programming Feature Selection (QPFS) algorithm [33], a multivariate filter technique that takes into account redundancy among features and relevance of each feature with respect to the task to perform. This algorithm is especially suitable for the problem here considered given its computational efficiency and outstanding performance in highly redundant databases. In particular, the QPFS feature selection method in a M-dimensional space consists of minimizing a multivariate quadratic function subjected to linear constraints as follows

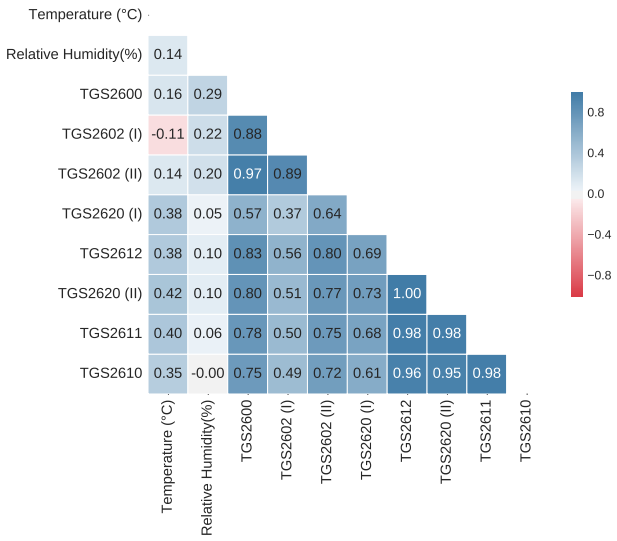


Fig. 2. Correlation between pairs of sensors computed with the complete time series in the dataset from chemical gas sensor array in turbulent wind tunnel [8].

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}(1-\alpha)\,\boldsymbol{w}^T Q \boldsymbol{w} - \alpha \boldsymbol{F}^T \boldsymbol{w} \tag{2}$$

$$\text{s.t.} \quad w_i \geq 0 \;\text{ for }\; i = 1, 2, \ldots, M\,, \tag{3}$$

$$\|\boldsymbol{w}\|_1 = 1\,, \tag{4}$$

where $\boldsymbol{w}$ is an M-dimensional vector, Q is a symmetric positive semidefinite matrix in $\mathbb{R}^{M\times M}$ with non-negative entries, and $\boldsymbol{F}$ is a vector in $\mathbb{R}^M$ with non-negative entries. Q represents the similarity among variables (redundancy), and $\boldsymbol{F}$ measures the similarity of the features with the target (relevance). The components of the solution vector $\boldsymbol{w}$ represent the normalized positive weight of each feature. Thus, the goal of Eqs. 2-4 is to select those features that provide a good trade-off between relevance and redundancy. Finally, the real parameter $\alpha \in [0, 1]$ enables to overweight the linear and the quadratic term in Eq. 2. In other words, α regulates the trade-off between relevance and redundancy. As an example, Pashami et al. adapt the QPFS algorithm to select a subset of sensors for detecting changes in the activity of a distant gas source from the response of an array of metal-oxide gas sensors deployed in an open sampling system. They use the Pearson correlation to quantify redundancy among features (matrix Q), and they use a measure based on the Fisher Index to determine the relevance of each sensor (vector $\boldsymbol{F}$). Their results show that selecting sensors with QPFS allows obtaining detection rates comparable with those corresponding to the best single sensor, while providing lower detection delays than the single sensor. For more details, the reader is referred to [29], [33].

IV. Sensor Calibration as a Supervised Learning Problem

Enose technology offers immense possibilities for environmental monitoring applications, but its proper calibration is crucial to guarantee the applicability of the sensing technology to the task that it is designed for. As stated in Section I, gas sensor calibration can be easily identified with classification and regression problems in machine learning, and during these years a large variety of calibration techniques has been investigated for chemical detection systems, including artificial neural networks, linear discriminants, multilayer perceptrons, k-NN classifiers, partial least square regressors, and more recently, Support Vector Machines [15], [34]–[37]. This section will review the classification/regression models and results in sensor calibration for the four enose applications described in Section I. The regression problem associated with the prediction of the number of people in a NASA spacecraft cabin simulator (Application III.1.) is addressed by applying the well-known Support Vector Regression (SVR) model with a RBF kernel [38]. The multiclass classification problems in Table I use Inhibitory Support Vector Machines (ISVMs) [19] and λ-Support Vector Machines, an extension of ISVMs. ISVMs have shown good performance in calibrating sensor arrays [6], [7], [9], [39] and their goal is to provide a simple algorithm for multiclass classification by directly integrating the concept of inhibition into the SVM formalism. However, the use of the ISVM classifier in sensor calibration settings, such as the one described in Algorithm 1, arises a fundamental question of whether the successive inclusion of training points leads to the optimal classifier. This point implies guaranteeing the Bayes consistency of the ISVM model, which can be addressed by analyzing the pointwise Fisher consistency (or classification calibration) of the classifier as this property states necessary and sufficient conditions to have Bayes consistency when a classifier minimizes a surrogate loss function [40]. ISVM authors show the consistency of the ISVM model for problems with two or three classes [19], such as the

active sensor calibration procedure (Application I.1) and the classification calibration problem for binary gas mixtures (Application II.1). However, when working with more than three classes, the ISVM model cannot guarantee the Bayes consistency of the classifier. That is why the λ-SVM model [7], an extension of ISVM, is proposed as a universally pointwise Fisher consistent multiclass classifier. The λ-SVM model is defined by a real parameter λ representing the margin of the positive points of a given class. The margin is set to 1 for points belonging to other classes. The ISVM classifier is a particular case of λ-SVM by setting $\lambda = 1$. Formally, given a training set of N patterns, $\{\boldsymbol{x_i}\}_{i=1}^{N}$, in which each point $\boldsymbol{x_i}$ belongs to a known class $\hat{y}_i \in [1, L]_{\mathbb{N}}$, the λ-SVM objective function is defined as follows,

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\sum_{j=1}^{L}\eta_{ij} \tag{5}$$

$$\text{s.t.} \quad \eta_{ij} \geq 0 \tag{6}$$

$$-1 - (\lambda - 1)\frac{y_{ij}+1}{2} + f_j(\mathbf{x}_i)y_{ij} + \eta_{ij} \geq 0\ , \tag{7}$$

where $\boldsymbol{w}$ is the concatenation of the hyperplanes of each class, $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L]$, $\{\eta_{ij}\}$ are the slack variables that provide room to handle the noisy data, and y_{ij} takes the value 1 if the pattern $\boldsymbol{x_i}$ belongs to class j (i.e. $\hat{y}_i = j$) and -1 otherwise. The cost parameter $C \in [0, \infty)$ establishes a trade-off between the two objectives of the model: maximizing the margin while classifying correctly as many training patterns as possible. The key difference between λ-SVMs and standard SVMs relies on λ-SVM's decision function, which includes an inhibitory term regulated by a scalar parameter μ. λ-SVM's decision function associated with the j-th class and the input pattern $\boldsymbol{x}_i$ is defined as $f_j(\boldsymbol{x_i}) = \langle \boldsymbol{w_j}, \Phi(\boldsymbol{x_i})\rangle - \mu\sum_{k=1}^{L}\langle \boldsymbol{w_k}, \Phi(\boldsymbol{x_i})\rangle$. It can be shown that the optimal value for μ is $\frac{1}{L}$, which can be directly obtained by minimizing the Lagrangian in Eqs. 5-7 [7], [19]. The classification of a data point $\tilde{\boldsymbol{x}}$ is determined by the maximum of the evaluation function for each class: $y(\tilde{\boldsymbol{x}}) = \arg\max_j f_j(\tilde{\boldsymbol{x}})$. Finally, the function Φ is a map from the input space to a higher dimensional space where the optimal hyperplanes, $\boldsymbol{w}_i$, are calculated. In the results here presented, the mapping function Φ is that associated with the RBF kernel with compact support. It can be shown that the multiclass classification function defined in Eqs. 5-7 is pointwise Fisher consistent for $\lambda \in (-\infty, 0) \cup ((L-2)/2, (L-1))$ [7]. This is the first multiclass hinge-loss function capable of giving unlimited weight to the positive examples without breaking the classification calibration property, which is shown to be beneficial in terms of classification accuracy and training times. In fact, the optimal λ value for the sensors calibration problem (Application I.2) is obtained in a classification calibrated scenario with $\lambda = 10,000$. For more detailed description and analysis of the λ-SVM method, the reader is referred to [7] and references therein.

A summary of data characteristics and sensor calibration results is shown in Table I. Column *Feature Type* indicates whether the input space for the classification/regression model is defined either by the time series captured by the sensors or by an aggregate of synthetic features (see Section III). *Column C/R* shows if the sensor calibration problem is a classification problem (C) or a regression (R) problem. In parenthesis, the number of classes in the classification problem is shown. Column *ML Model* indicates the machine learning model used for sensor calibration. Column *Err.* shows the classification error in the case of classification problems, and the mean relative error for the regression problem. Column *Ref.* lists the papers related to the corresponding dataset and application, in which a detailed description of the experimental setup for each application can be found. According to the results shown in Table I, machine learning techniques applied to electronic nose data analysis allow the design and development of effective chemical detection systems.

V. Further Work

The works presented in this review show the capability of machine learning techniques to properly calibrate sensors and monitor environments, but they are mainly obtained in controlled environments where data are correctly labelled. Future lines of research in machine learning applied to electronic nose data analysis should be focused on extending existing algorithms or designing new ones to overcome some of the challenges arising from the deployment of sensor arrays in open and non-controlled environments. This scenario brings up several challenges for the machine learning community as many of the most popular machine learning algorithms cannot be directly applied. Some of these difficulties include: (i) high correlation between gas signals and environmental variables such as humidity and temperature, which makes it difficult to extract the chemical information captured by the sensors; (ii) lack of labeled data or the presence of noise in the labels, which makes it difficult to adopt machine learning strategies without the supervision of an expert; (iii) existence of drift when sensors are working for long periods of time, which compromises the applicability of the machine learning algorithms that assume distribution stationarity.

Acknowledgment

IR-L acknowledges partial support by Spain's grants TIN2013-42351-P (MINECO), S2013/ICE-2845 CASI-CAM-CM (Comunidad de Madrid), and TIN2015-70308-REDT. IR-L and RH acknowledge the partial support by 3ª Convocatoria de Proyectos de Cooperacion Interuniversitaria UAM–Banco Santander EEUU.

References

[1] M. Ogawa and T. Togawa, "Monitoring daily activities and behaviors at home by using brief sensors," in *Microtechnologies in Medicine and Biology, 1st Annual International, Conference On. 2000.* IEEE, 2000, pp. 611–614.

[2] T. Oyabu, A. Okada, O. Manninen, and D.-D. Lee, "Proposition of a survey device with odor sensors for an elderly person," *Sensors and Actuators B: Chemical*, vol. 96, no. 1, pp. 239–244, 2003.

TABLE I
DESCRIPTION AND RESULTS OF MACHINE LEARNING MODELS FOR ELECTRONIC NOSE APPLICATIONS.

Application	#samples	#attributes	Feature Type	C/R	ML Model	Err.	Ref.
Active sensor calibration (Application I.1.)	1,800	128	Synthetic features	C(3)	ISVM	0.50%	[5], [6]
Sensor calibration (Application I.2.)	13,910	128	Synthetic features	C(6)	λ-SVM	0.35%	[5], [7]
Sensor calibration (Application II.1.)	180	23,688	Time series	C(2)	ISVM	2.11%	[8], [9]
NASA human activity monitoring (Application III.1.)	672	4,320	Time series	R	SVR	0.32	[10]

[3] K. Danzer and L. A. Currie, "Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998)," *Pure and Applied Chemistry*, vol. 70, no. 4, pp. 993–1014, 1998.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[5] J. Fonollosa, I. Rodríguez-Luján, and R. Huerta, "Chemical gas sensor array dataset," *Data in brief*, vol. 3, pp. 85–89, 2015.

[6] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, and R. Huerta, "On the calibration of sensor arrays for pattern recognition using the minimal number of experiments," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 123–134, 2014.

[7] I. Rodriguez-Lujan, R. Huerta *et al.*, "A Fisher consistent multiclass loss function with variable margin on positive examples," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 2255–2292, 2015.

[8] J. Fonollosa, I. Rodríguez-Luján, M. Trincavelli, and R. Huerta, "Dataset from chemical gas sensor array in turbulent wind tunnel," *Data in brief*, vol. 3, pp. 169–174, 2015.

[9] J. Fonollosa, I. Rodríguez-Luján, M. Trincavelli, A. Vergara, and R. Huerta, "Chemical discrimination in turbulent gas mixtures with mox sensors validated by gas chromatography-mass spectrometry," *Sensors*, vol. 14, no. 10, pp. 19 336–19 353, 2014.

[10] J. Fonollosa, I. Rodriguez-Lujan, A. V. Shevade, M. L. Homer, M. A. Ryan, and R. Huerta, "Human activity monitoring using gas sensor arrays," *Sensors and Actuators B: Chemical*, vol. 199, pp. 398–402, 2014.

[11] C. D. Natale, S. Marco, F. Davide, and A. D'Amico, "Sensor-array calibration time reduction by dynamic modelling," *Sens. Actuators, B.*, vol. 25, no. 1–3, pp. 578–583, 1995.

[12] M. L. Homer, A. V. Shevade, L. Lara, R. Huerta, A. Vergara, and M. K. Muezzinoglu, "Rapid analysis, self-calibrating array for air monitoring," in *42nd International Conference on Environmental Systems*, 2012, p. 3457.

[13] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, and R. Huerta, "Chemical gas sensor drift compensation using classifier ensembles," *Sensors and Actuators B: Chemical*, vol. 166, pp. 320–329, 2012.

[14] M. Urízar Salinas *et al.*, "Low cost artificial noses with communication and collaboration capabilities," 2014.

[15] R. Gutierrez-Osuna and A. Hierlemann, "Adaptive microsensor systems," *Annual Review of Analytical Chemistry*, vol. 3, pp. 255–276, 2010.

[16] F. Herrero-Carrón, D. J. Yáñez, F. de Borja Rodríguez, and P. Varona, "An active, inverse temperature modulation strategy for single sensor odorant classification," *Sensors and Actuators B: Chemical*, vol. 206, pp. 555–563, 2015.

[17] R. Lomasky, "Active Acquisition of Informative Training Data," Ph.D. dissertation, TUFTS UNIVERSITY, 2009.

[18] R. Lomasky, C. E. Brodley, M. Aernecke, D. Walt, and M. Friedl, "Active class selection," in *Machine learning: ECML 2007*. Springer, 2007, pp. 640–647.

[19] R. Huerta, S. Vembu, J. M. Amigó, T. Nowotny, and C. Elkan, "Inhibition in multiclass classification," *Neural computation*, vol. 24, no. 9, pp. 2473–2507, 2012.

[20] R. M. Castro and R. D. Nowak, "Minimax Bounds for Active Learning," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2339–2353, 2008.

[21] N. Barsan, D. Koziej, and U. Weimar, "Metal oxide-based gas sensor research: How to?" *Sensors and Actuators B: Chemical*, vol. 121, no. 1, pp. 18–35, 2007.

[22] R. Huerta, S. Vembu, M. K. Muezzinoglu, and A. Vergara, *Dynamical svm for time series classification*. Springer, 2012.

[23] M. K. Muezzinoglu, A. Vergara, R. Huerta, N. Rulkov, M. I. Rabinovich, A. Selverston, and H. D. Abarbanel, "Acceleration of chemo-sensory information processing using transient features," *Sensors and Actuators B: Chemical*, vol. 137, no. 2, pp. 507–512, 2009.

[24] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: a review," *Sensors Journal, IEEE*, vol. 2, no. 3, pp. 189–202, 2002.

[25] J. Fonollosa, A. Vergara, and R. Huerta, "Algorithmic mitigation of sensor failure: Is sensor replacement really necessary?" *Sensors and Actuators B: Chemical*, vol. 183, pp. 211–221, 2013.

[26] E. Martinelli, G. Magna, A. Vergara, and C. Di Natale, "Cooperative classifiers for reconfigurable sensor arrays," *Sensors and Actuators B: Chemical*, vol. 199, pp. 83–92, 2014.

[27] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, and A. Perera, "Drift compensation of gas sensor array data by common principal component analysis," *Sensors and Actuators B: Chemical*, vol. 146, no. 2, pp. 460–465, 2010.

[28] E. Phaisangittisagul, H. T. Nagle, and V. Areekul, "Intelligent method for sensor subset selection for machine olfaction," *Sensors and Actuators B: Chemical*, vol. 145, no. 1, pp. 507–515, 2010.

[29] S. Pashami, A. J. Lilienthal, and M. Trincavelli, "Detecting changes of a distant gas source with an array of MOX gas sensors," *Sensors*, vol. 12, no. 12, pp. 16 404–16 419, 2012.

[30] A. Vergara and E. Llobet, "Sensor selection and chemo-sensory optimization: toward an adaptable chemo-sensory system," *Bioinspired solutions to the challenges of chemical sensing*, p. 143, 2012.

[31] X. R. Wang, J. T. Lizier, T. Nowotny, A. Z. Berna, M. Prokopenko, and S. C. Trowell, "Feature selection for chemical sensor arrays using mutual information," *PloS one*, vol. 9, no. 3, p. e89840, 2014.

[32] X. R. Wang, J. T. Lizier, A. Z. Berna, F. G. Bravo, and S. C. Trowell, "Human breath-print identification by E-nose, using information-theoretic feature selection prior to classification," *Sensors and Actuators B: Chemical*, vol. 217, pp. 165–174, 2015.

[33] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, "Quadratic programming feature selection," *The Journal of Machine Learning Research*, vol. 11, pp. 1491–1516, 2010.

[34] A. Hierlemann and R. Gutierrez-Osuna, "Higher-order chemical sensing," *Chemical reviews*, vol. 108, no. 2, pp. 563–613, 2008.

[35] S. Marco and A. Gutiérrez-Gálvez, "Signal and data processing for machine olfaction and chemical sensing: a review," *Sensors Journal, IEEE*, vol. 12, no. 11, pp. 3189–3214, 2012.

[36] M. K. Muezzinoglu, A. Vergara, and R. Huerta, "A unified framework for Volatile Organic Compound classification and regression," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*. IEEE, 2010, pp. 1–7.

[37] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, "Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring," *Sensors and Actuators B: Chemical*, vol. 215, pp. 618–629, 2015.

[38] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[39] A. Vergara, J. Fonollosa, J. Mahiques, M. Trincavelli, N. Rulkov, and R. Huerta, "On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines," *Sensors and Actuators B: Chemical*, vol. 185, pp. 462–477, 2013.

[40] A. Tewari and P. L. Bartlett, "On the consistency of multiclass classification methods," *The Journal of Machine Learning Research*, vol. 8, pp. 1007–1025, 2007.