Chemometrics and Intelligent Laboratory Systems 130 (2014) 123–134

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

On the calibration of sensor arrays for pattern recognition using the minimal number of experiments

Irene Rodriguez-Lujan [a,*], Jordi Fonollosa [a], Alexander Vergara [b], Margie Homer [c], Ramon Huerta [a]

[a] BioCircuits Institute, University of California, La Jolla, San Diego, CA 92093, USA
[b] Biomolecular Measurement Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8362, USA
[c] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

ARTICLE INFO

Article history:
Received 21 June 2013
Received in revised form 18 October 2013
Accepted 19 October 2013
Available online 26 October 2013

Keywords:
Electronic nose
Gas sensing
Support Vector Machines
Fast calibration
Pattern recognition
Gas discrimination
Active learning

ABSTRACT

We investigate optimal experiment selection to train a classifier on gas sensor arrays to get the maximal possible performance in a limited number of experiments. In gas sensing, while collecting data for a particular sensor array, one has to choose what gas and concentration level is going to be presented in the next experiment. It is an active decision by the operator selecting the experiments and training the classifiers. Can the algorithm be trained sooner rather than later? Can we minimize the costs of collecting the data in terms of the man-hour of the operator and the expenses of the experiment itself? Active control sampling provides a way to deal with the challenge of minimizing the calibration costs and is applicable to any situation where experimental selection is parameterized by an external control variable. Our results indicate that active sampling strategies can only improve a random selection of experiments over a wide range of concentration of gasses. However, random or uninformed selection is fairly close. Additionally, our active sampling methodology reveals that, when there is no prior knowledge about the range of concentrations to which the sensor will be exposed during real operation, sensor must be calibrated over the entire working range, not just high concentrations. In fact, our results show that it is especially important to include low concentrations in the calibration since the lack of these values would dramatically decrease the performance of the system.

1. Introduction

The calibration of gas sensor arrays is an expensive, but necessary, process to establish the functional relationship between measured values and analytical quantities. Traditionally, calibration includes first, the selection of the functional form of a computational model; second, the estimation of the corresponding model parameters and the errors based on a training dataset; and third, the model validation [1]. The resulting computational model is then used to deconvolve new measurements and predict the analyte amount/class. However, after a certain period of time, the performance of the model degrades due to the changing characteristics of the sensing elements, and the system needs to be recalibrated.

In particular, the calibration (or recalibration) of solid state gas sensor arrays has been investigated for decades using nonlinear multivariate techniques [2,3]. During these years a large variety of calibration techniques has been investigated for chemical detection systems, including, but not limited to, artificial neural networks, linear discriminants, multilayer perceptrons, k-NN classifiers, partial least square regressors, and more recently, Support Vector Machines [4–7]. However, irrespective of the selected data processing technique, a training dataset needs to be collected to perform the calibration of an analytical system. The generation of the training dataset represents a significant cost in terms of time and budget due to the expenses of the experiment itself and the dedication of technicians. This situation is especially critical in applications where the acquisition of new samples is costly, such as air quality control in space ships, environmental monitoring of public spaces, and industrial leak detection, among others. Additionally, systems based on Metal Oxide (MOX) gas sensors, which still are a common choice for chemical detection applications due to their sensitivity, low cost, ease of operation, ability to detect large number of chemicals, and robustness [8,9], are dynamic systems that show a transient response when exposed to a constant stimulus [10]. Hence, in order to collect a thorough training dataset, it is necessary to capture the complete transient response of the sensors for each training example, thereby making the calibration of a MOX gas sensor array an extensively expensive, laborious, and time consuming operation [11,12].

In order to reduce the frequency of the recalibrations and the associated costs, methodologies aimed at extending the time between recalibrations have been presented that attenuate the effects of sensor drift [10,13–15], sensor failure [16–18], or sensor poisoning [19,20]. However, the selection of the training examples utilized to reduce the cost of the calibration process has been overlooked. The practitioner

* Corresponding author at: BioCircuits Institute, University of California, San Diego, 9500 Gilman Dr., Mail Code 0402, La Jolla, CA 92093-0402, USA. Tel.: +1 858 534 6876.
E-mail addresses: irenerodriguez@ucsd.edu (I. Rodriguez-Lujan), fonollosa@ucsd.edu (J. Fonollosa), alexander.vergaratinoco@nist.gov, vergara@ucsd.edu (A. Vergara), margie.l.homer@jpl.nasa.gov (M. Homer), rhuerta@ucsd.edu (R. Huerta).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.10.012

has to design an experimental protocol to collect the training dataset to maximize the accuracy after the calibration, while reducing the number of training examples and calibration costs.

The challenge of the calibration process is to sample the space of admissible control parameters to achieve superior and stable performance after the operation. When performing a system calibration, each new training example (measurement) has an associated class label (gas type) and a control parameter (gas concentration). It is generally believed that gasses presented at higher concentrations are easier to classify than those at lower concentrations. However, when a chemical detection system is deployed in its operating environment, it is responsible for identifying various gasses from a wide range of concentrations. When selecting a new training example, the question that the experimentalist faces is: what is the next training example (gas class and concentration) that maximizes learning?

This idea of selecting intelligently the next experiment to calibrate the sensor array is in line with the active sampling algorithms proposed in machine learning. The goal of active learning is to improve the performance of a preexisting classifier on-the-fly while accelerating the learning process by incorporating and labeling samples to the learning process, taking into account the information provided by the previous examples. While most theories and methods in machine learning assume independent and identically distributed observations, they make use of *passive learning* strategies based on non-adaptive (usually uniform) sampling. However, there are several theoretical and empirical works in the literature demonstrating the superiority of active sampling over passive learning in terms of generalization error, uncertainty, and stability [21–29]. Unfortunately, the application of the proposed criteria to determine the suitability of active sampling for a real-world problem is unfeasible due to their strong assumptions about the distribution of the data and the noise.

Different paradigms can be found in the active sampling literature namely, *adaptive sampling* or *query learning* [30], *instance selection* or *selective sampling* [31,32] and *pool-based* learning [33]. Adaptive sampling assumes that the learning algorithm actively creates or selects unlabeled samples to be labeled by an expert. An alternative to query learning is selective sampling that, under the hypothesis that obtaining unlabeled instances is inexpensive, decides whether or not to request the label for samples drawn according to the data distribution. Finally, pool-based learning assumes that there exists a big pool of unlabeled instances while there is a small subset of labeled data. Then, the sampling strategy chooses the best query after ranking all the instances in the pool according to a certain measure. Thus, the main difference between pool-sampling and selective sampling is that the former needs to evaluate all the instances in the pool while the latter considers instances individually. For a more detailed review on active sampling, the reader is referred to Ref. [21]. Therefore, the active control sampling strategy suggested in this work is in line with the adaptive sampling approach as the learning algorithm creates queries by selecting the gas concentration level (control parameter) from a set of feasible values.

In contrast to active sensing methodologies that actively adapt the sensor properties or exposure conditions to optimize the system performance [34–36], in this paper we propose a methodology to select the best training examples to calibrate the system without modifying the configuration of the sensor array. Our methodology will allow us to determine (i) whether an active sampling strategy outperforms a uniform selection of experiments; and (ii) the range of concentrations to be used to calibrate the sensor array, especially when there is no prior knowledge about the range of concentrations to which the sensor will be exposed during real operation. The approach is based on Support Vector Machines, a technique that was successfully introduced to classify e-nose data [7,37–39] with actual measurement recordings from a 16-element MOX gas sensor array.

The paper is organized as follows: In Section 2 we describe the experimental setup and dataset. Then, we introduce the formal description of the problem in Section 3, followed by the results and discussion (Section 4) and the conclusions of this work (Section 5).

2. Experimental details

2.1. Dataset and measurement collection procedure

We implemented the suggested active sampling methodology to a dataset recorded over 1 month utilizing sixteen screen-printed MOX gas sensors commercialized by Figaro Inc. [40].[1] The resulting dataset comprises 1800 recordings 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 distribution of the recordings as a function of the gas type and its concentration is given in Fig. 1. It shows that the distributions for the three gasses are similar, with the concentrations of ethanol being slightly biased toward the lower values. Therefore, the acquired dataset is suitable to explore the calibration strategies since it is well balanced for all the classes and concentrations.[2]

In order to visualize the distribution of the data, Fig. 2 shows the data projected into the 2-dimensional and 3-dimensional spaces defined by the first two and three principal components, respectively, and obtained from Principal Component Analysis using the correlation matrix [41]. The first principal component retains 63.59% of the variance, while second and third components represent 16.50% and 11.51% of the variance, respectively. The three principal components represent 91.60% of the total variance of the data but they do not provide a good representation to clearly classify the three gasses.

As previously described in [10], to construct our dataset, we placed the sensor array – a set of chemical sensors tagged by the manufacturer as TGS2600, TGS2602, TGS2610, and TGS2620 (four of each) – into a 60 ml volume polytetrafluoroethylene/stainless steel air-tight test chamber, a vapor flow cell into which the gaseous substances are injected at a constant flow and in a random order. The test chamber is attached in series to a vapor delivery system that provides the selected concentrations of the chemical substances by means of three digital mass flow controllers (Bronkhorst High-Tech B.V. [42]) and the calibrated gas cylinders (Air gas [43]). The entire measurement system setup is fully operated by a computerized environment and provides versatility for setting the concentrations with high accuracy and in a highly reproducible manner (see Fig. 3).

To generate the dataset, we adopted a measurement procedure consisting of the following three steps. First, in order to stabilize the sensors and measure the baseline of the sensor response, we circulated synthetic dry air (10% R.H. measured at 25 ± 1 °C) through the sensing chamber during 50 s. Second, we randomly added one of the analytes of interest to the carrier gas (in our case synthetic dry air) and made it circulate through the sensor chamber during 100 s. Finally, we re-circulated clean dry air for the subsequent 200 s to clear up the sensors and the test chamber to have the system ready for a new measurement. The measurements were performed at a constant flow rate of 200 ml/min (sccm).

To capture the sensors' responses and control their operating temperature (indirectly controlled by the sensor operating voltage), we adapted a PC platform fitted with the appropriate data acquisition and serial communication cards with a National Instruments data acquisition board controlled by a customized LabVIEW code [44]. The dynamic response of each sensor was recorded at a sample rate of 100 Hz while the gas sensor array was kept at a stable operating temperature (400 °C).[3]

[1] Certain commercial equipment, instruments or materials are identified in this report to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

[2] The complete acquired dataset is freely available at the UCI repository at http://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations.

[3] We do not have access to the actual sensing surface temperature due to packaging, but a look-up table relating it to the voltage applied to the sensor's embedded heating element can be found upon request in Ref. [40]. Note that the effect of varying the sensors' operating temperature was investigated in a previous work, by following information theoretic optimization formalisms [45].

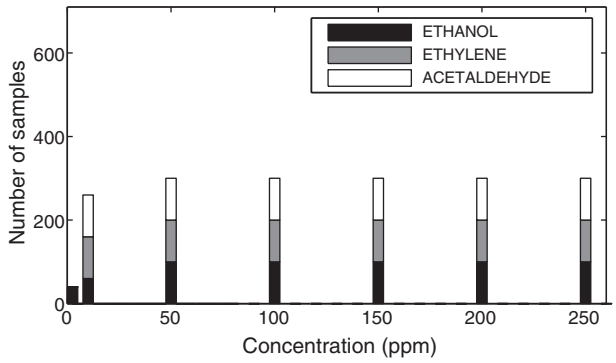

Fig. 1. The acquired dataset is balanced for the three classes and concentrations to explore calibration strategies without bias.

2.2. Data processing and feature extraction

MOX gas sensors typically respond with a monotonically smooth change in the conductance of the sensing layer due to the adsorption/desorption reaction processes of the exposed chemical analyte substance. For a more detailed explanation of the operating principle of this popular sensing technology the reader is referred to reviews in Ref. [8,46–48]. In our experimental setup the active layers (i.e., sensor types) and operating temperature of the sensors are fixed. Therefore, the interaction processes between the sensor and the analyte identity and/or concentration dosage are the only two factors that, as a pair, shape the response profile, defining thereby the identity of the chemical analyte of interest.

The sensor response is read in the form of the resistance across the active layer of each sensor; hence, each measurement produced a 16-channel time series. We represented each time series with an aggregate of eight features reflecting the sensor response. In particular, we considered two distinct types of features in the creation of this dataset: two steady-state features and six features reflecting the sensor dynamics.

First, the so-called steady-state feature (ΔR) is defined as the difference of the maximal resistance and the baseline,

$$\Delta R = \max_k r[k] - \min_k r[k], \tag{1}$$

and its normalized version is expressed by the ratio of the maximal resistance and the baseline values,

$$||\Delta R|| = \frac{max_k r[k] - min_k r[k]}{min_k r[k]}, \tag{2}$$

where $r[k]$ is the sensor resistance time profile, k is the discrete time indexing in the recording interval $[0,T]$, and T is the duration of the measurement.

Second, we extracted an aggregate of features based on the exponential moving average (ema_α) to reflect the sensor dynamics of the increasing/decaying transient portion of the sensor responses [48]. The ema_α transform evaluates the rising/decaying portions of the sensor resistance by considering the maximum/minimum values of $y[k]$ of the following first-order digital filter:

$$y[k] = (1-\alpha)y[k-1] + \alpha(r[k]-r[k-1]), \tag{3}$$

with the initial condition $y[0] = 0$, and where α is the smoothing parameter of the filter varying from 0 to 1. Since different values of α provide different feature values and different information of the transient response, we computed the ema_α filter for three values of $\alpha = \{0.1, 0.01, 0.001\}$ for both the rising and the decaying stages. Therefore, each of the 16 sensors used in the study contributes with 8 features, thereby yielding a 128 element feature vector per measurement.

3. Problem definition

Considering a class of problems in which new data can only be obtained by laboratory experiments (not numerical simulations) and assuming that all experiments share a common control parameter that may or may not correlate with the final classification results, our goal is to determine whether, given a limited budget of time and/or cost, an optimal strategy for experiment selection exists. In other words, we want to calibrate the sensor array as fast and efficiently as possible. More precisely, assuming that we are able to successfully complete batches of experiments of a predefined size, we want to optimize the values of the generic control parameter to be used in the immediately following batch without wasting the preceding experiments and without knowing beforehand the final number of experiments that we will be able to carry out.

For our specific application, we would like to determine if a strategy for experiment selection based on active sampling is more effective than an uninformative sampling procedure to calibrate a sensor array used to discriminate three different gasses. This problem can be reformulated as deciding 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. In particular for our case, the control parameter is the gas concentration c. Since classification performance is more heavily impacted by the selection of the concentration level than the selection of the label of the targeted gas [49,50], we chose the manipulation of the external concentration as the control parameter. The system to calibrate can be formulated from the point of view of machine learning as a multi-class classification task with three

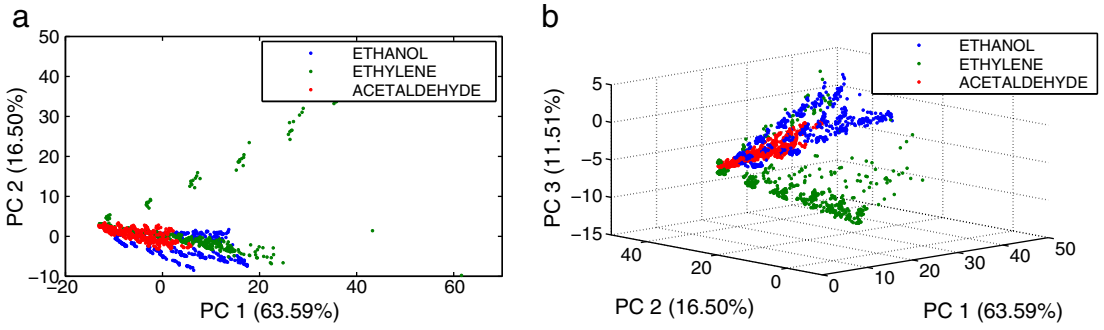

Fig. 2. (a) Projection of the dataset onto the first two principal components. (b) Projection of the dataset onto the first three principal components.

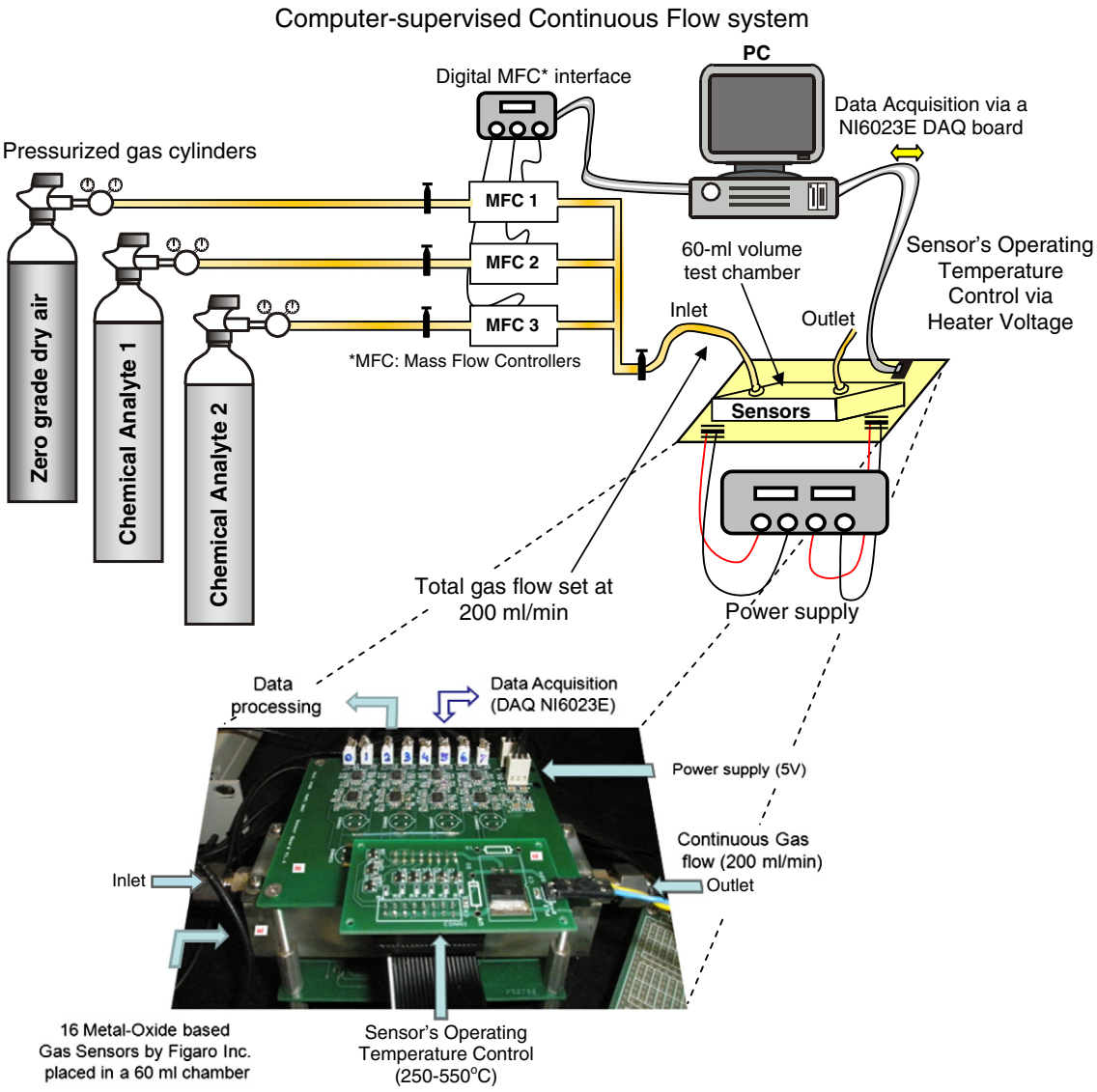

Fig. 3. Experimental setup used for data acquisition. The sensor responses are recorded in the presence of the analyte in gaseous form diluted at different concentrations in dry air. The measurement system operates under a fully computerized environment with minimal human intervention, which accurately controls the gas concentrations to the sensing chamber while keeping constant the total flow. Therefore, no changes in the flow or flow dynamics are reflected in the sensor response, (i.e., only the presence of an odorant will be reflected in the sensor response). Moreover, since the system is continuously supplying gas to the sensing chamber (either clean dry air or a chemical component), the amount of gas molecules in the sensing chamber is homogeneously distributed.

classes (gasses ethanol, ethylene and acetaldehyde). Each of the experimental recordings formed by 128 features (see Section 2.2) represents a pattern for the classifier. Note that the concentration level of each recording is not explicitly provided to the classifier.

3.1. Inhibitory Support Vector Machine

In this work, we used an Inhibitory Support Vector Machine (ISVM) with RBF kernel as multi-class classifier [39] to classify ethanol, ethylene and acetaldehyde as a function of the multivariate response of the sensor array. The choice of ISVM was motivated by its consistency for three-class classification problems and its robustness on small datasets, such as the type used for sensor calibration.

ISVMs are motivated by Support Vector Machines, which have become one of the most popular classifiers nowadays due to their excellent performance in many domains besides their solid mathematical basis. Intuitively, SVMs establishes a trade-off between maximizing the generalization capability of the model and classifying correctly as many training patterns as possible. SVMs were initially proposed to solve binary classification problems (two classes) [51], and ISVM is an extension of SVM to provide a simple algorithm for multiclass classification by directly integrating the concept of inhibition into the SVM formalism. The objective of the inhibition mechanism behind the ISVM algorithm is to find a hyperplane associated with each class that exerts downward pressure on the rest hyperplanes while trying to maximize its generalization capability. For more details, the reader is referred to [39].

Formally, given a training set of N patterns, $\chi = \{x_i\}_{i=1}^{N}$, in which each point x_i belongs to a known class $\hat{y}_i \in [1, L]_{\mathbb{N}}$, the ISVM objective function is defined as follows,

$$\begin{aligned} &min_w \quad E(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\sum_{j=1}^{L}\eta_{ij} \\ &\text{s.t.} \quad \eta_{ij} \geq 0 \\ &\qquad y_{ij} f_j(x_i) - 1 + \eta_{ij} \geq 0. \end{aligned} \tag{4}$$

In our case, each pattern x_i represents the multivariate response of the sensor array when exposed to a pure substance y_i (ethanol, ethylene or acetaldehyde) and thus, $L = 3$. The cost parameter $C \in [0, \infty)$ establishes a trade-off between the two objectives of the model: maximizing the margin and classifying correctly as many training patterns as possible. $\{\eta_{ij}\}$ are the slack variables which provide room to handle the noisy data (note that L slack variables are associated with each data point, corresponding to the hyperplanes $\{w_j\}_{j=1}^{L}$). y_{ij} takes the value 1 if the pattern x_i belongs to class j (i.e. $\dot{y}_i = j$) and -1 otherwise. Finally, $f_j(x_i)$ is the ISVM evaluation function for the class j and the data point x_i and it is defined as:

$$f_j(x_i) = \left\langle w_j, \Phi(x_i) \right\rangle - \frac{1}{L}\sum_{k=1}^{L}\langle w_k, \Phi(x_i)\rangle. \tag{5}$$

Φ is a mapping function from the original input space to a higher-dimensional space (feature space) where the optimal hyperplane is calculated. Therefore, a linear classifier in the higher-dimensional space is equivalent to a non-linear classifier in the original space if the mapping function Φ is non-linear. Interestingly, the SVM formulation does not require to compute the mapping function Φ explicitly since the optimization problem in Eq. (4) can be expressed as a function of inner products in the feature space via the kernel trick [52]. Thus, only the inner product between any two points in the feature space, x and x', needs to be defined as a kernel function $K(x,x')$. In our experiments, we utilized a RBF kernel for its ability to deal with complex decision surfaces:

$$\mathrm{RBF}\left(x, x^{'}\right) = exp\left(-\frac{\gamma \| x - x^{'} \|^2}{M}\right), \tag{6}$$

M being the total number of features, and $\gamma \in [0, \infty)$ being a parameter inversely proportional to the kernel width.

Finally, the classification of a data point $\tilde{x}$, $y(\tilde{x})$, is determined by the maximum of the evaluation functions:

$$y(\tilde{x}) = arg\max_j f_j(\tilde{x}). \tag{7}$$

3.2. Active sampling

As stated above, our first goal is to determine if a strategy for experiment selection based on active control sampling is more effective than uninformative sampling to calibrate a sensor array when there exists a limited budget of time and/or cost. This limitation in the number of recordings affects the performance of machine learning algorithms, generally more effective on large datasets, and thus, a protocol for experiment selection is critically important. In addition, given the unpredictable setbacks or equipment malfunctions occurring during the experiments, it is very difficult to know the final number of successful measurements beforehand.

In our particular application, we want to effectively discriminate among three types of gasses. As classification performance is more heavily impacted by concentration selection than gas label selection [49,50], we will focus our work on the manipulation of the external control parameter disregarding the information of the label. At this point it is worth mentioning that the uniform distribution of the conditional probabilities $P(gas|c)$ in our dataset (Fig. 1) ensures uninformative (random) sampling on the label. Moreover, the uniform distribution of the data with respect to the concentration c makes the results unbiased [53].

To test an active sampling strategy, we simulate a cross-validation technique [41] in which we draw the concentration c according to certain probability distribution. We work at batch level (of size B), and we assume that the sampling distribution in each batch is constant. Then, given the best sampling strategy in the preceding batches, we want to compute the best sampling strategy for the following batch in order to effectively calibrate the sensor array.

We assume a canonical sampling probability distribution over the space of gas concentrations $p(c)$. Note that our motivation in this paper is not to find the optimal distribution to sampling concentrations c, but rather to analyze the benefits of sampling with a general distribution of concentrations. We set $p^{\kappa}(c) \propto exp(-\kappa c)$, being κ the rate of the sampling distribution. This probability distribution not only enables us to tilt the sampling distribution toward higher or lower concentrations, but also to recover the uniform distribution. More precisely, higher concentrations are sampled more frequently when $\kappa < 0$, lower concentrations are sampled more frequently when $\kappa > 0$, and uniform sampling is recovered when $\kappa = 0$.

Initially, the best κ value for the first batch is that with the lowest cross-validation error using an ISVM as classifier. In the next batches, we follow a greedy approach: given the sequence $\{\kappa_1, \ldots, \kappa_m\}$ corresponding to the best κ values for the previous m batches, the best κ value for the $(m+1)$-th batch, κ_{m+1}, will be that with the lowest ISVM cross-validation error in the $(m+1)$-th batch when the classifier is trained using not only the experiments in the $(m+1)$-th batch but also the experiments already selected in the preceding batches. In other words, assuming that we have already sampled the first, second, and mth batches by following the distributions suggested by κ_1, κ_2 and κ_m, respectively, we now want then to optimize the value of κ for the following experimental batch. Fig. 4 illustrates this approach in an example with batches of size $B = 10$. The optimal strategy for the first batch (experiments $x_1, x_2, \ldots, x_{10}$) is drawing the concentrations with a sampling rate κ_1 (in the example $\kappa_1 < 0$, which favors higher concentrations). Assuming that no information is available on previous measurements before the acquisition of the first batch, the optimal κ_1 is simply selected as the one that minimizes the cross-validation error. In the second batch (experiments $x_{11}, x_{12}, \ldots, x_{20}$), we need to establish the sampling rate κ_2 that minimizes the cross-validation error in this batch when the classifier is trained with the experiments $x_1, x_2, \ldots, x_{10}$ already sampled according to κ_1 and the experiments $x_{11}, x_{12}, \ldots, x_{20}$ to be sampled according to κ_2. In the example shown in Fig. 4, the optimal value for the second batch is $\kappa_2 > 0$, which is biased toward the lower concentrations. The same procedure follows for the third batch, yielding as a result that the best sampling rate for x_{21} to x_{30} is $\kappa_3 = 0$ (uniform, not biased toward any range of concentrations). Assuming that we have enough time and budget to carry out the next batch of experiments, we are at the point to determine the best sampling rate for the fourth batch without wasting $x_1, \ldots, x_{30}$.

Though our practical approach to the problem is greedy in the sense that we locally optimize the performance of the system for the immediately following batch, it is possible to have a global view of the sampling distribution up to the m-th batch (i.e. for the first $m \times B$ experiments). The probability of sampling at concentration c in the m-th batch (we will refer to this probability as *the joint sampling distribution at batch m*) is the sum of the probabilities of having sampled at concentration c in any of the m batches.[4] Thus, the joint probability at batch m and at concentration c provided by $\kappa_1, \kappa_2, \ldots, \kappa_m$ is given by the following equation:

$$p_m(c|\kappa_1, \kappa_2, \ldots, \kappa_m) = \frac{1}{m}\sum_{i=1}^{m} \frac{exp(-\kappa_i c)}{\sum_{c^{'} \in C} exp(-\kappa_i c^{'})}, \tag{8}$$

[4] Note that the sampling distributions in each batch are independent.

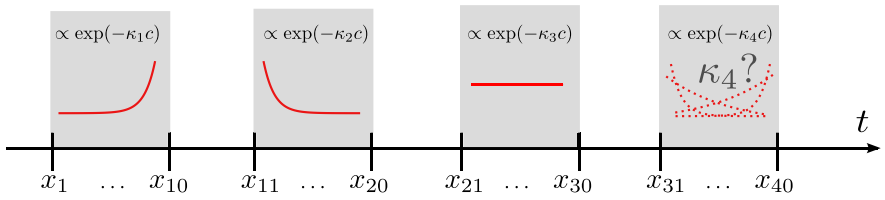

Fig. 4. Illustration of our experimental design. Each x_i represents one recording. In this case, knowing the best κ values for the three first batches, we want to determine that value of κ in the fourth batch which minimizes the ISVM cross-validation error.

where 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 $\sum_{c' \in C} p(c) = 1$. Finally, the factor $\frac{1}{m}$ comes from the normalization of the joint probability $p_m(c|\kappa_1, \kappa_2, \ldots, \kappa_m)$ so that $\sum_{c' \in C} p_m\left(c'|\kappa_1, \kappa_2, \ldots, \kappa_m\right) = 1$.

Fig. 5 shows an example of the sampling distribution for each batch and the corresponding joint distribution for a sequence of three batches. This case can be identified with the example shown in Fig. 4, in which higher values are favored in the first batch ($\kappa_1 = -0.03$), the sampling distribution is biased toward the lower concentrations in the second batch ($\kappa_2 = 0.03$), and there is no bias in the sampling strategy corresponding to the third batch ($\kappa_3 = 0$). Since the first batch is sampled with $\kappa_1 = -0.03$, higher values of concentrations are favored. Note that for the first batch, no previous measurements are available; therefore, the joint sampling distribution and the sampling distribution in the batch are coincident. As the opposite scenario is presented in the second batch, where the sampling strategy is biased toward the lower concentrations, the joint action of the first two batches leads to a joint distribution favoring both lower and higher concentrations. Finally, bias is absent in the sampling strategy corresponding to the third batch ($\kappa_3 = 0$), where the joint distribution is less biased toward lower and higher concentrations.

4. Results and discussion

The aim of the following experiments is twofold: first, to determine if an active sampling strategy is more efficient to calibrate a chemical system than an uninformative approach; and second, to analyze the classification performance of the sensor array as a function of the range of concentrations presented during the calibration process.

4.1. Active sampling versus non-informative strategy

To test if the active sampling strategy described in Section 3.2 sufficiently improves calibration, we compared the ISVM cross-validation error of our dynamic approach against the one corresponding to an uninformative strategy in which the concentrations are randomly sampled. In all the following experiments, we used the experimental dataset described in Section 2 formed by 1800 128-dimensional patterns.

The details of the algorithm used to implement the active sampling strategy are given in Algorithm 1. The grid of κ values ranged from −0.03 to 0.03 with a stepwise resolution of 0.001. The uninformative sampling procedure can be easily recovered from Algorithm 1 by imposing $\kappa = 0$ for all the batches.

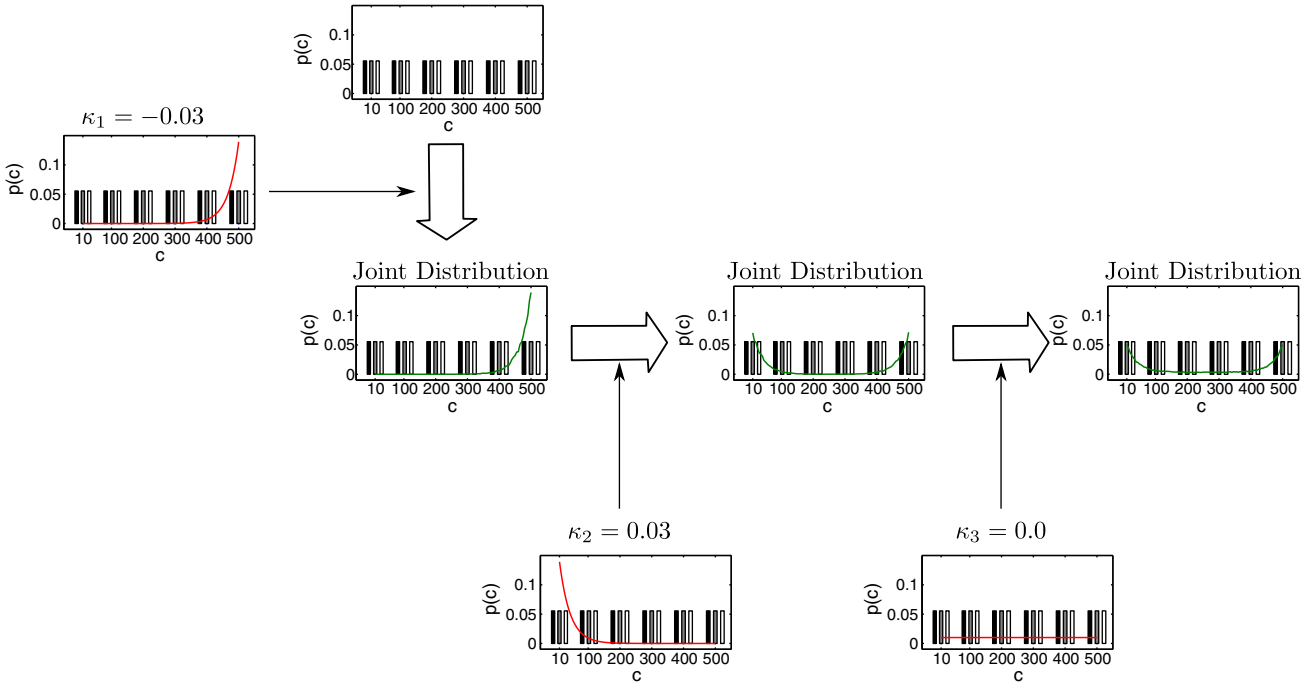

Fig. 5. Example of the relationship between the sampling distribution in each batch and the joint sampling distribution in an active sampling strategy for three consecutive batches.

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 we used $N_{cv} = 100$.

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 $t \leftarrow 1 \ldots N_{CV}$
      $S \leftarrow$ sample $(m+1) \times B$ points according to Eq. (8).
      $T \leftarrow U \setminus S$
      Train the ISVM classifier on S
      $v_t \leftarrow$ test error on T
    end for
    $\widetilde{\mu}_{C,\gamma} \leftarrow$ mean($\mathbf{v}$)
    $\widetilde{\sigma}_{C,\gamma} \leftarrow$ std($\mathbf{v}$)
  end for
end for
$\mu \leftarrow \min_{C,\gamma} \widetilde{\mu}_{C,\gamma}$
$\sigma \leftarrow \min_{C,\gamma} \widetilde{\sigma}_{C,\gamma}$

Fig. 6 shows the error (logarithmic scale) when training with 1 to 10 batches, each one of size 20. This means that the total number of training samples ranges from 20 to 200 and the testing set size varies from 1780 to 1600. Assuming that no equipment malfunctions during the experiment procedure and that each measurement takes 5 min to complete, it could take anywhere from 100 min to 16 h to collect the data used in these experiments. Once the training size (the number of batches) is specified, we explore now the classification performance over a range of κ s. Several conclusions can be derived from this figure: (i) the overall performance of the model significantly improves when more samples are available: the error rate varies from 12.48% in the first batch to 0.38% in the last batch. That is, with 200 recordings (or even less), the ISVM algorithm is able to almost perfectly discriminate the three gasses; (ii) it seems that the sampling strategy strongly affects the performance of the classifier in the first stages while it is not so determinant when enough samples have been presented to the sensor array. In particular, the first two batches clearly show that κ values close to 0 yield the best results and batches 3 and 4 still present a slight improvement when κ is close to 0. However, when more information becomes available (batches 5–10), there is not a predominant strategy as the error rate has already converged.

Fig. 7 summarizes the results presented in Fig. 6 and compares them against the non-informative/random strategy. Fig. 7(a) shows the lowest error rate for each batch (logarithmic scale) and Fig. 7(b) represents the difference between the error rates of the non-informative sampling and the active sampling strategies. Note that the error rate of the active sampling strategy is always lower than that of the uniform sampling (Fig. 7(b)), though the difference between both approaches in terms of classification accuracy is small. In order to objectively compare the classification performance of both approaches, we evaluated the statistical significance of their error rate differences by applying the one-tailed paired t-test for equal means to the two sets of error rates, one set for uninformative sampling and one set for the active sampling strategy. The test was one-tailed because the null hypothesis is that the uniform sampling strategy is as good as or better than the active sampling or vice versa. The test was also paired because we measured the performance of both approaches over the same test partition. To do so, we randomly selected 1000 samples from the entire dataset and kept them as the test set. A percentage of the remaining 800 recordings was drawn according to either the non-informative strategy or the active sampling procedure. Both ISVMs were trained with the optimal cross-validation parameters and the resulting models were evaluated over the test set of 1000 recordings. This procedure was repeated 100 times thus, obtaining 100 paired measures of the performance. The results of the t-test (at the 5% significance level) as a function of the number of batches revealed that the active sampling is not statistically significantly better than the uniform/non-informative sampling. In summary, the

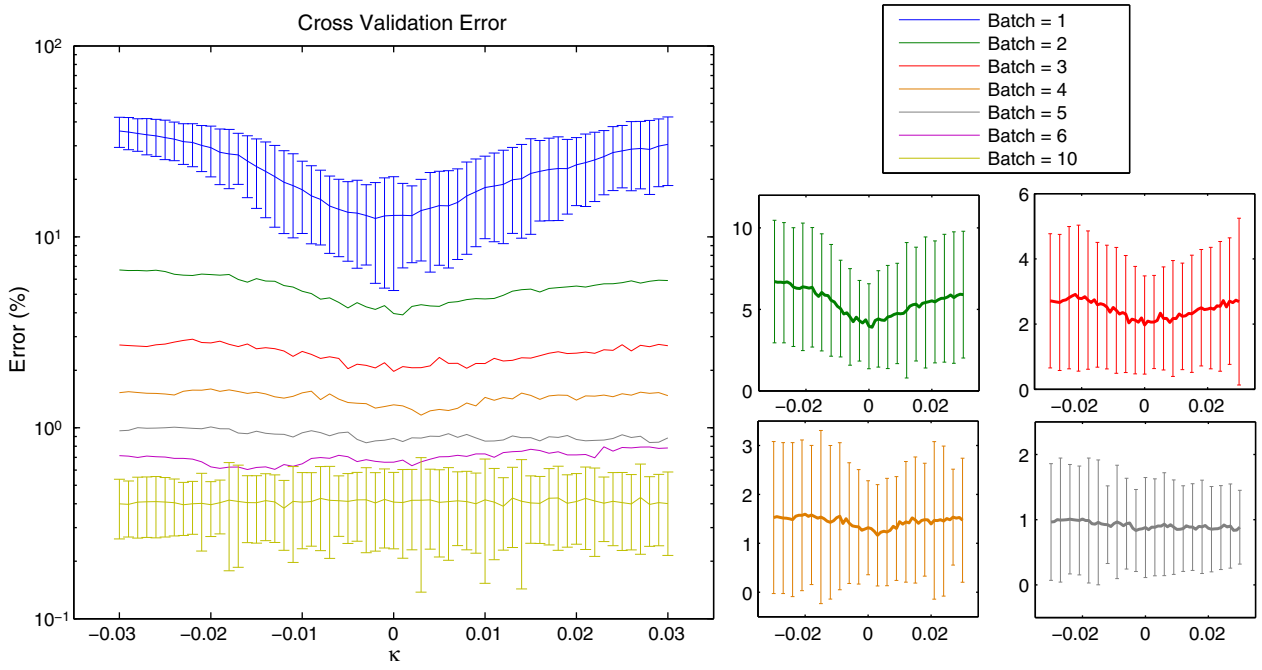

Fig. 6. Cross-validated error (log scale) using Algorithm 1 and all the examples in the database for increasing values of the rate κ. The main plot shows the cross-validated error (logarithmic scale) for different number of batches ranging from 1 to 10. To improve the quality of the figure, only the error bars for 1 batch and 10 batches are shown. The right-hand plots show the cross-validated error (linear scale) for the second (top-left), third (top-right), fourth (bottom-left) and fifth (bottom-right) batches. Note that while in the first batches a clear advantage is obtained from sampling with κ rates close to zero, the sampling strategy is irrelevant from the fourth batch since the error rate has already converged.

active sampling approach can only improve a random selection of experiments.

Despite the similar behavior of both approaches in terms of classification accuracy, the active sampling strategy is visibly different from the uninformative one as shown in Fig. 8(a), which presents the evolution of the sampling distribution along the 10 batches. Each subfigure shows the optimal sampling distribution restricted to the samples in the current batch (*Batch distribution*) as well as the joint distribution computed according to Eq. (8). It can be noticed that the active sampling strategy begins sampling the higher concentrations ($\kappa_1 = -0.0020$), then it moves toward the lower concentrations ($\kappa_2 = 0.0010$), finally, in the third batch, the best choice is the uniform sampling ($\kappa_3 = 0.0$). In the following batches, the optimal κ value oscillates between negative and positive values, i.e. the active sampling distribution in each batch alternately favors lower or higher concentrations. Nevertheless, these pronounced oscillations occurring at batch level become smoother when the joint distribution is considered. In fact, the joint distribution is almost uniform in some batches (4, 5, 6, 9 and 10). A quantification of the bias of the joint distribution toward the lower or higher concentrations is provided in Fig. 8(b), where the ratio between the probability of sampling the lowest concentrations over the probability of sampling the highest concentrations according to the joint distribution is shown. In Fig. 8(b), a ratio equal to 1 corresponds to the uniform samplings as it equalizes the probabilities of lower and higher concentrations while the values of the ratio much larger or smaller than 1 indicate a significant bias toward the lower or higher concentrations, respectively. The results presented in Fig. 8(b) show a low-high ratio for active sampling very close to 1, implying that a wide spectrum of concentrations is sampled. In fact, the largest difference between the sampling probabilities of lower and higher concentrations amounts to 0.6, which implies that the higher concentrations are sampled 1.67 times more frequently than the lower ones but the chances of sampling at low concentration levels are still significant.

Though the differences between the active sampling and non-informative strategies are noticeable, they coincide in choosing a wide range of concentrations to calibrate the sensor array. The main difference between both approaches relies on the initial batches, in which the active sampling algorithm tends to favor the higher concentrations. However, both alternatives seem to be plausible solutions for the first stages of the learning process: while the uniform sampling can be justified by the lack of information available to the classifier, the preference for the higher concentrations of the active sampling approach can be explained by the presence of *evident* discriminant information due to the larger signal-to-noise-ratio.

4.2. Optimal concentration range

Our second pertaining question is how to select the best range of concentrations to calibrate a sensor array. Is it really needed to use a broad range of concentrations or is it possible to restrict the range of concentrations while still guaranteeing a good performance for a broad range of concentrations? In order to determine which situation holds in our calibration problem, we performed the analysis by considering all the possible concentration pairs (c_{min}, c_{max}) to calibrate the sensor array. Concentration levels 2.5 μmol/mol and 10 μmol/mol were grouped to have a uniform distribution of gasses at each concentration level. Fig. 9 shows the error rates obtained when the model is evaluated either inside or outside the range of concentrations. In both cases, the performance is evaluated using the ISVM parameters C and γ with the lowest cross-validated error.

Several interesting conclusions can be drawn from Fig. 9. First, Fig. 9(a) shows the generalization ability of the sensor array to classify unseen/unknown concentration levels. The best performance was obtained in the upper left corner, which means that concentrations ranging from the lowest to relatively high values (larger than 100 μmol/mol) yield the best generalization results. What this result in fact means is that it is crucial to calibrate the sensor array in a concentration range including low and high concentration levels: the exclusion of the lowest concentrations (2.5 and 10 μmol/mol) clearly deteriorates the performance of the system, while restricting the calibration only to the lowest concentrations is not good either.

Second, the low error rates in Fig. 9(b) with respect to those in Fig. 9(a) reveal that for a optimal calibration the system should be trained in the same range of concentrations at which it will be exposed when operating autonomously. Additionally, Fig. 9(b) shows that narrow intervals present the lowest cross-validation errors, but they provide the worst generalization results for unseen concentrations (Fig. 9(a)). In fact, the intervals that are more affected by this phenomenon are located in the diagonal (where the intervals are constrained to a single value) with exception of the bottom-left element that corresponds to the lowest concentration interval (2.5 and 10 μmol/mol). It is likely that the poor generalization of these narrow intervals is due to overfitting: the sensor is extremely accurate for the range of concentrations at which it has been trained but it is vague for concentrations out of range. However, this is not the case for the lowest concentrations (bottom-left case) presumably because of their low signal-to-noise ratio. It is also remarkable that the increasing accuracy of both Fig. 9(a) and Fig. 9(b) with the number of training samples matches with the behavior of most machine learning algorithms, generally more effective with large datasets.

In order to clearly differentiate and quantify the absolute performance in Fig. 9, Table 1 shows the test and cross-validation errors (in percentage) for all the concentration pairs [c_{min},c_{max}] and using 200 recordings to calibrate the sensor array. Notice that for the interval [10,250] the test error does not exist according to our experimental setup. In all cases the cross-validation errors are significantly lower than their corresponding test errors, which confirms the fact that the optimal experimental design lies in calibrating the sensor array at the range of concentrations to which it will be exposed. Table 1 also confirms that (i) the wider the range of concentrations, the better generalization performance for unseen concentrations; (ii) the narrower the interval of concentrations, the lower cross-validation error; and (iii) the lower the concentrations, the more challenging to discriminate and thus, their inclusion in the calibration process is crucial. For example, according to the results reported in Table 1, calibrating the sensor in the range [10,200] yields a classification error of 2.26% when tested for concentrations out of range; i.e. 250 μmol/mol. On the contrary, if calibration is carried out considering concentrations in the interval [50,250], the classification error rate for recordings at concentration level of 10 μmol/mol dramatically increases to 26.13%.

Our proposed active learning methodology presented in Section 3.2 is sufficiently general to permit the use of any classification algorithm as well as sensing technology to select the calibration examples without showing qualitatively any detrimental effect in the final results. On the one hand, although we tested our methodology with MOX gas sensors, it can in principle be implemented on any type of sensing technology that employ simple and robust transduction schemes with partial specificity or cross-reactive selectivity to a variety of chemical compounds. For the particular case of the MOX gas sensors, the Clifford–Tuma model [54,55] relates the measured sensor resistance with the analyte concentration, showing that the sensor responses are monotonic when the operating temperature is fixed. In other words, the amplitude of the sensor response when the sensor is exposed to an analyte compared to the sensor response when it is exposed to clean air (no analyte exposure) is larger as higher concentration levels of the analyte are induced. In fact, the monotonic response is a desired property for any kind of sensor type since it makes the sensor response an injective function and, ideally, the gas concentration can be predicted certainly from the amplitude of the sensor response. Most of the sensor technologies are monotonic, as supported by the corresponding sensing principles that include conductivity variation of chemiresistive micro-films [54,55], radiation absorption in optical spectroscopy

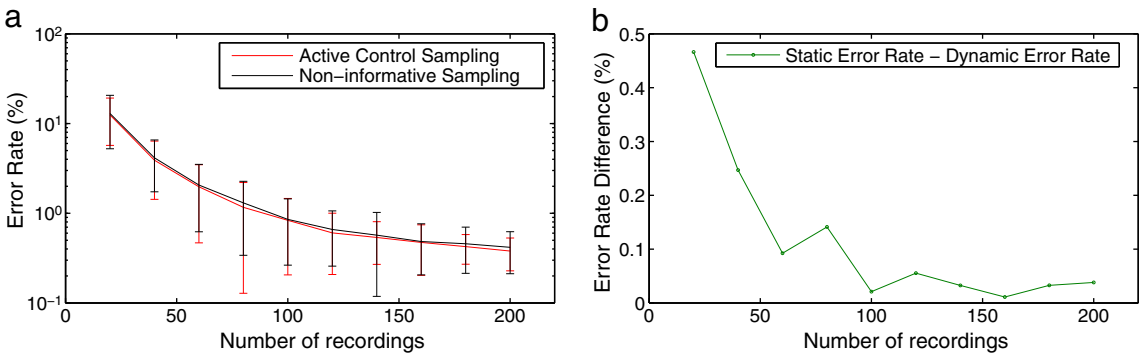

Fig. 7. (a) Optimal errors with their standard deviation as a function of the number of recordings (equivalently, batches) for the active sampling and the non-informative strategies. (b) Error rate difference between the active (dynamic) and non-informative (static) sampling strategies.

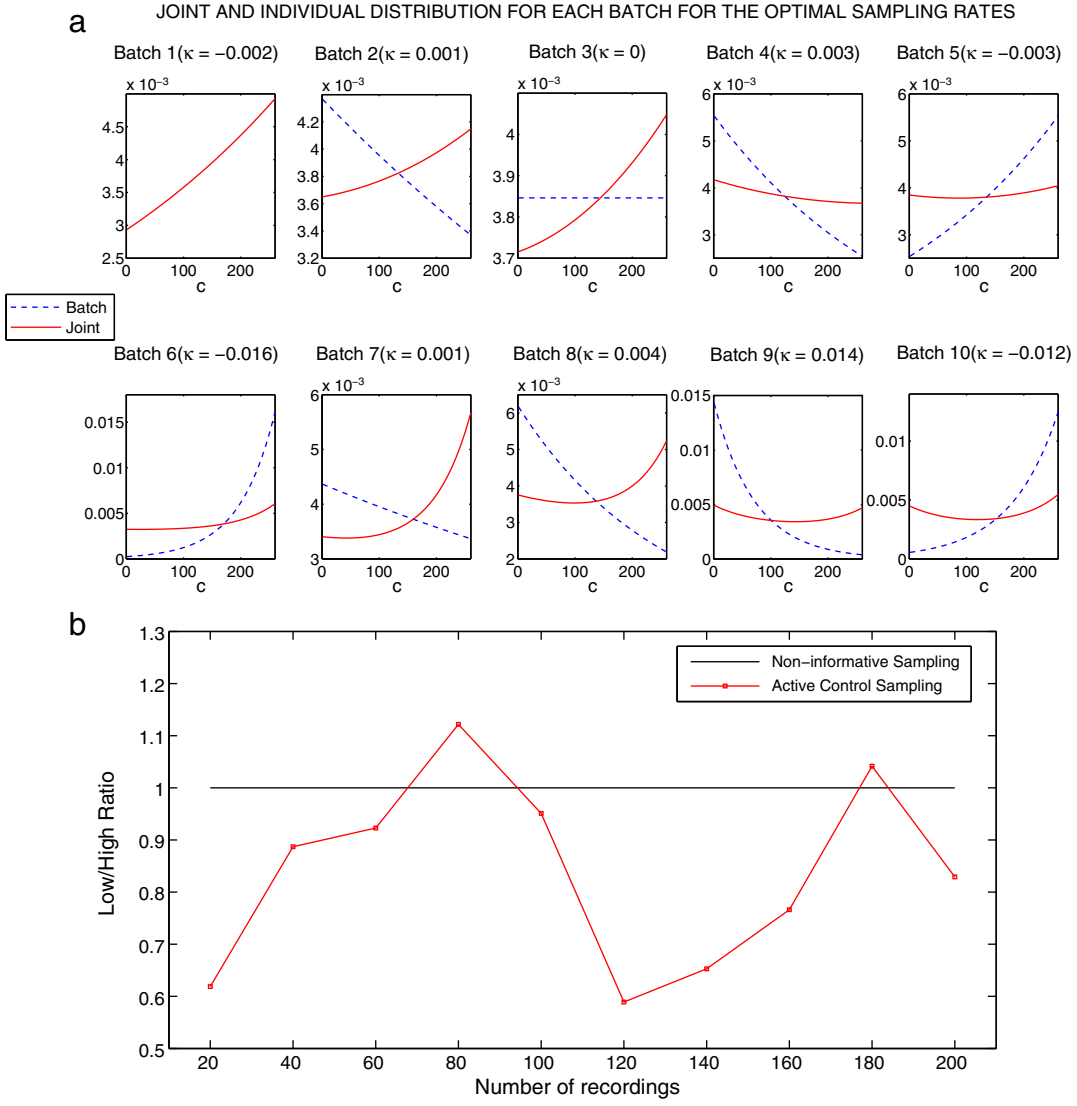

Fig. 8. (a) Optimal sampling distribution for each batch and joint sampling distribution up to the current batch obtained by our active sampling strategy. Note that while the joint distribution is biased toward the higher concentrations in the three first batches, it becomes almost uniform for batches 4–6. In batches 7 and 8, the joint distribution starts to favor the higher concentrations but note that, at this point, the error rate has already converged and thus, there is not a strategy that significantly improves the calibration. (b) Ratio between the probability of sampling the lowest concentrations over the probability of sampling the highest concentrations for the active sampling strategy and the non-informative strategy (ratio constantly equal to 1). In the case of the active sampling strategy, the ratio of the joint sampling strategy considering the optimal κ values for each batch is shown. This ratio gives an idea of how much the sampling distribution is far from being uniform. That is, values close to 1 mean that extreme concentrations are sampling with similar probabilities while values significantly smaller/larger than 1 indicate that the sampling distribution is highly skewed toward the highest/lowest concentrations. These results indicate that the active sampling distribution does not clearly favor the lower or higher concentrations but it maintains all the time a representative proportion of all the concentrations.

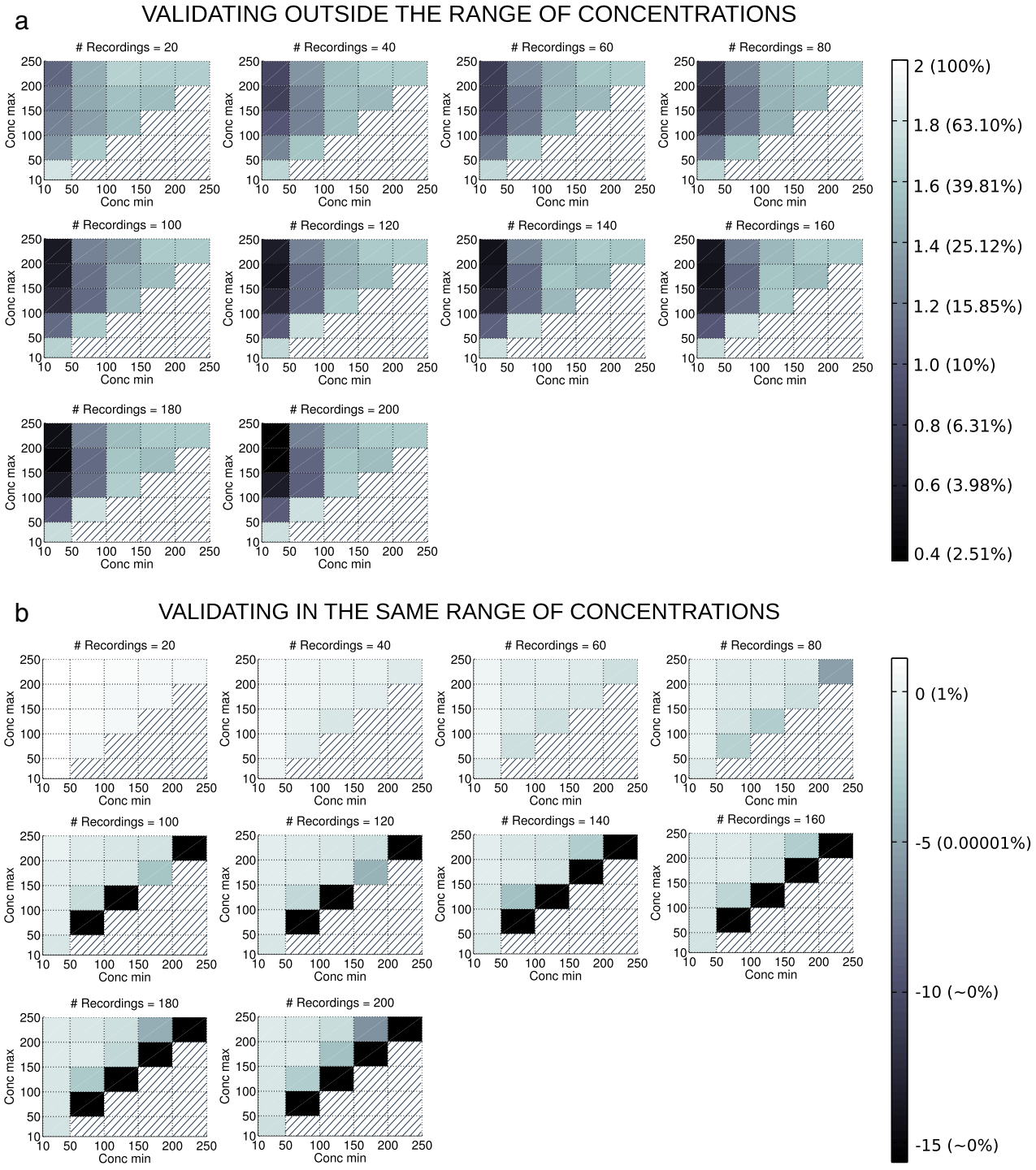

Fig. 9. Classification error rates using Algorithm 1 with uniform sampling and calibrating the sensor array with different intervals of concentrations and number of recordings. For each interval of concentrations, $[c_{min},c_{max}]$, a train set was generated exclusively with those recordings in this range of concentrations (dataset U in Algorithm 1). (a) Error rate (logarithmic scale) obtained in the evaluation of recordings outside the range of concentrations at which the sensor has been calibrated. The lowest errors are obtained in the upper left corner which means that concentrations ranging from the lowest to relatively high values (larger than 100 μmol/mol) yield the best generalization results. (b) Error rate (logarithmic scale) obtained in the evaluation of recordings on the same concentration levels at which the sensor has been calibrated. The intervals with the lowest cross-validation error are those with a narrow range of concentrations (elements in and around the diagonal) but they are the intervals that provide the worst generalization results for unknown concentrations.

[56], frequency shift of surface acoustic wave sensors [57], current produced in amperometric electrochemical sensors [58], temperature variation in catalytic sensors [59], primary spectral wavelength shift in optically based vapor sensors [60], among others [61]. Therefore, since different technologies show monotonic sensor responses, one can expect that the proposed algorithm will behave similarly when changing the sensor technology.

On the other hand, Algorithm 1 can be adapted to any classification or regression function by simply replacing the ISVM classifier by an arbitrary classification or regression model (see [7] for example). In our case, the ISVM model is especially suitable for our particular task given its good performance with a limited number of examples (first batches) and its competitiveness in larger datasets [39]. Nevertheless, qualitatively similar results are expected when applying other pattern

Table 1
Average error rate (%) in the classification of the three gasses when the sensor is calibrated using 200 recordings with concentrations uniformly distributed in the interval $[c_{min},c_{max}]$ and tested for concentrations out of the range $[c_{min},c_{max}]$. The cross-validated error obtained by evaluating the sensor in the same spectrum of concentrations at which it has been calibrated is also given in parentheses. The best result in each scenario is marked in bold. As already discussed in Fig. 9, observe how the lowest error rate for evaluations out of the range of concentrations is obtained when the sensor is calibrated over the widest range of concentrations. On the contrary, the lowest error rate for evaluations in the same range of concentrations corresponds to models calibrated using a small spectrum of concentrations; that is, results in or around the diagonal.

c_{min} / c_{max}	10	50	100	150	200	250
250	–	26.13 ± 0.59	39.77 ± 0.61	53.35 ± 0.15	52.30 ± 0.07	47.85 ± 0.03
	(0.42 ± 0.02)	(0.21 ± 0.01)	(0.09 ± 0.02)	(0.01 ± 0.01)	$\mathbf{(0.00 \pm 0.00)}$	$\mathbf{(0.00 \pm 0.00)}$
200	$\mathbf{2.26 \pm 0.17}$	16.66 ± 0.05	31.36 ± 0.39	38.91 ± 0.04	39.42 ± 0.01	–
	(0.28 ± 0.02)	(0.18 ± 0.01)	(0.02 ± 0.01)	$\mathbf{(0.00 \pm 0.00)}$	$\mathbf{(0.00 \pm 0.00)}$	–
150	2.23 ± 0.08	11.62 ± 0.05	37.55 ± 0.28	32.98 ± 0.07	–	–
	(0.30 ± 0.01)	(0.20 ± 0.01)	$\mathbf{(0.00 \pm 0.00)}$	$\mathbf{(0.00 \pm 0.00)}$	–	–
100	3.83 ± 0.13	10.92 ± 0.09	43.48 ± 0.16	–	–	–
	(0.13 ± 0.01)	$\mathbf{(0.00 \pm 0.00)}$	$\mathbf{(0.00 \pm 0.00)}$	–	–	–
50	10.45 ± 0.24	57.46 ± 0.37	–	–	–	–
	(0.16 ± 0.01)	$\mathbf{(0.00 \pm 0.00)}$	–	–	–	–
10	58.34 ± 0.73	–	–	–	–	–
	(0.05 ± 0.02)	–	–	–	–	–

recognition algorithms to our gas classification problem for several reasons. First, the active learning results are qualitatively independent of the classifier. One can expect that any classifier will obtain the major advantage with respect to the non-informative sampling in the first batches (training examples will contribute more significantly to the system calibration when little information is available). Additionally, given that the procedure described in Algorithm 1 recovers the uniform distribution when $\kappa = 0$ and can be applied with any classifier, our methodology can only improve the performance of the non-informative approach. Second, other classification methods may change the percentage of accuracy by a few tenths of a percent, but the final results should not change qualitatively since some of the most popular machine learning algorithms are closely related. For example, Adaboost, Support Vector Machines, Neural Networks, Least Squares Support Vector Machines or Logistic Regression can be formulated as Large Margin algorithms [62,63] and Support Vector Machines are related to probabilistic exponential models [64,65,39]. These equivalences go even beyond the pattern recognition problem per se [66]. The common statistical framework under these algorithms makes that, under certain assumptions, they converge to classification models with certain statistical properties [67,68]. Finally, as discussed in Section 4.2, the results on the analysis of the range of concentrations to calibrate the sensor array are supported by some principles of statistical learning theory [69] independent of the underlying machine learning algorithm.

In short, our formulation above can in principle be implemented on any type of pattern recognition technique as well as sensing technology based on any transduction principle that provides cross sensitivities to a set of analytes.

5. Conclusions

We have analyzed the performance of a strategy for experiment selection based on active sampling to calibrate a gas sensor array to classify three gasses under limited budgetary time resources and/or costs that makes fast calibration almost mandatory. Our control parameter is the concentration at which the gasses are presented to the sensor. We have defined an experimental protocol based on a sampling distribution for the targeted gas concentrations to determine if an active/informed strategy is more effective to calibrate the sensor array than the uninformative experiment selection. We have tested both approaches over our extensive dataset of 1800 gas recordings of three gasses. Our results and analysis indicate that:

- An active sampling strategy can only improve the non-informative/random sampling approach.
- Constraining the sampling distribution in the early stages of calibration is likely to generate an inaccurate classifier.
- The classifier needs to have access to the same range of concentrations as in the real operating environment because the best performance is obtained when the range of concentrations for calibration and testing coincide.
- When there is no prior knowledge about the operating concentrations, it is especially important to include low concentrations in the calibration since the lack of these values dramatically decreases the performance of the system.
- The wider the range of concentrations to calibrate the sensor array, the better generalization performance for unseen (out of range) concentrations.
- When enough number of samples are presented, the sensor array is able to discriminate between three pure gaseous substances (ethanol, ethylene and acetaldehyde) with a error rate below 0.5% using active sampling and uninformative selection of gasses.

Acknowledgments

This work has been supported by the Jet Propulsion Laboratory under the contract number 2013-1479652 and partially financed by the North Atlantic Treaty Organization (NATO) under the Science for Peace & Security Program, grant no. SPS-984511. Alexander Vergara was financially supported by the NIST/NIH Research Associateship program administered by the National Research Council and partially financed by NATO. The authors also thank Joanna Zytkowicz for proofreading and revising the manuscript.

References

[1] K. Danzer, L. Currie, Guidelines for calibration in analytical chemistry, Pure Appl. Chem. 70 (4) (1998) 993–1014.
[2] W. Carey, S.S. Yee, Calibration of nonlinear solid-state sensor arrays using multivariate regression techniques, Sens. Actuators, B 9 (2) (1992) 113–122.
[3] S. Sekulic, M.B. Seasholtz, Z. Wang, B.R. Kowalski, S.E. Lee, B.R. Holt, Nonlinear multivariate calibration methods in analytical chemistry, Anal. Chem. 65 (19) (1993) 835A–845A.
[4] A. Hierlemann, R. Gutierrez-Osuna, Higher-order chemical sensing, Chem. Rev. 108 (2008) 563–613.
[5] R. Gutierrez-Osuna, A. Hierlemann, Adaptive microsensor systems, Annu. Rev. Anal. Chem. 3 (2010) 255–276.
[6] S. Marco, A. Gutierrez-Galvez, Signal and data processing for machine olfaction and chemical sensing: a review, IEEE Sens. J. 12 (11) (2012) 3189–3214.
[7] M.K. Muezzinoglu, A. Vergara, R. Huerta, A unified framework for volatile organic compound classification and regression, Neural Networks (IJCNN), The 2010 International Joint Conference on, IEEE, 2010, pp. 1–7.
[8] N. Barsan, D. Koziej, U. Weimar, Metal oxide-based gas sensor research: how to? Sens. Actuators, B 121 (1) (2007) 18–35.
[9] J. Fonollosa, L. Fernández, R. Huerta, A. Gutiérrez-Gálvez, S. Marco, Temperature optimization of metal oxide sensor arrays using mutual information, Sens. Actuators, B 187 (2013) 331–339.

[10] A. Vergara, S. Vembu, T. Ayhan, M.A. Ryan, M.L. Homer, R. Huerta, Chemical gas sensor drift compensation using classifier ensembles, Sens. Actuators, B (2012) 166–167320–329.
[11] C.D. Natale, S. Marco, F. Davide, A. D'Amico, Sensor-array calibration time reduction by dynamic modelling, Sens. Actuators, B 25 (1–3) (1995) 578–583.
[12] M.L. Homer, A.V. Shevade, L. Lara, R. Huerta, A. Vergara, M.K. Muezzinoglu, Rapid analysis, self-calibrating array for air monitoring, 42nd International Conference on Environmental Systems, 2012.
[13] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, S. Marco, Drift compensation of gas sensor array data by orthogonal signal correction, Chemom. Intell. Lab. Syst. 100 (1) (2010) 28–35.
[14] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, A. Perera, Drift compensation of gas sensor array data by common principal component analysis, Sens. Actuators, B 146 (2) (2010) 460–465.
[15] S.D. Carlo, M. Falasconi, E. Sánchez, A. Scionti, G. Squillero, A.P. Tonda, Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation, Pattern Recognit. Lett. 32 (13) (2011) 1594–1603.
[16] M. Pardo, G. Faglia, G. Sberveglieri, M. Corte, F. Masulli, M. Riani, Monitoring reliability of sensors in an array by neural networks, Sens. Actuators, B 67 (1–2) (2000) 128–133.
[17] J. Fonollosa, A. Vergara, R. Huerta, Sensor failure mitigation based on multiple kernels, Sensors 2012 IEEE, 2012, pp. 1–4.
[18] J. Fonollosa, A. Vergara, R. Huerta, Algorithmic mitigation of sensor failure: is sensor replacement really necessary? Sens. Actuators, B 183 (2013) 211–221.
[19] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, S. Marco, Poisoning fault diagnosis in chemical gas sensor arrays using multivariate statistical signal processing and structured residuals generation, Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on, 2007, pp. 1–6.
[20] P. Reimann, A. Dausend, A. Schutze, A self-monitoring and self-diagnosis strategy for semiconductor gas sensor systems, Sensors, 2008, IEEE, 2008, pp. 192–195.
[21] R.M. Castro, R.D. Nowak, Minimax bounds for active learning, IEEE Trans. Inf. Theory 54 (5) (2008) 2339–2353.
[22] D. MacKay, Information-based objective functions for active data selection, Neural Comput. 4 (4) (1992) 590–604.
[23] D.A. Cohn, Z. Ghahramani, M.I. Jordan, Active learning with statistical models, J. Artif. Intell. Res. 4 (1996) 129–145.
[24] Y. Freund, H.S. Seung, E. Shamir, N. Tishby, Selective sampling using the query by committee algorithm, Mach. Learn. 28 (2–3) (1997) 133–168.
[25] N. Cesa-Bianchi, A. Conconi, C. Gentile, Learning probabilistic linear-threshold classifiers via selective sampling, COLT, Vol. 2777 of Lecture Notes in Computer Science, Springer, 2003, pp. 373–387.
[26] G. Blanchard, D. German, Hierarchical testing designs for pattern recognition, Ann. Statistics 33 (3) (2005) 1155–1202.
[27] M. Saar-Tsechansky, F. Provost, Active sampling for class probability estimation and ranking, Mach. Learn. 54 (2) (2004) 153–178.
[28] N. Roy, A. Mccallum, Toward optimal active learning through sampling estimation of error reduction, In Proc. 18th International Conf. on Machine Learning (ICML), Morgan Kaufmann, 2001, pp. 441–448.
[29] T. Poggio, R. Rifkin, S. Mukherjee, P. Niyogi, General conditions for predictivity in learning theory, Nature 428 (6981) (2004) 419–422.
[30] D. Angluin, Queries and concept learning, Mach. Learn. 2 (4) (1987) 319–342.
[31] L.E. Atlas, D.A. Cohn, R.E. Ladner, Training connectionist networks with queries and selective sampling, NIPS, Morgan Kaufmann, 1989, pp. 566–573.
[32] D.A. Cohn, L.E. Atlas, R.E. Ladner, Improving generalization with active learning, Mach. Learn. 15 (2) (1994) 201–221.
[33] D.D. Lewis, W.A. Gale, A sequential algorithm for training text classifiers, SIGIR, ACM/Springer, 1994, pp. 3–12.
[34] R. Gosangi, R. Gutierrez-Osuna, Active temperature modulation of metal-oxide sensors for quantitative analysis of gas mixtures, Sens. Actuators, B 185 (2013) 201–210.
[35] T. Nakamoto, S. Ustumi, N. Yamashita, T. Moriizumi, Y. Sonoda, Active gas/odor sensing system using automatically controlled gas blender and numerical optimization technique, Sens. Actuators, B 20 (2–3) (1994) 131–137.
[36] T. Nakamoto, H. Matsushita, N. Okazaki, Improvement of optimization algorithm in active gas/odor sensing system, Sens. Actuators, A 50 (3) (1995) 191–196.
[37] M. Pardo, G. Sberveglieri, Classification of electronic nose data with support vector machines, Sens. Actuators, B 107 (2) (2005) 730–737.
[38] A. Shmilovici, G. Bakir, S. Marco, A. Perera, Finding the best calibration points for a gas sensor array with support vector regression, Intelligent Systems, 2004. Proceedings. 2004 2nd International IEEE Conference, vol. 1, IEEE, 2004, pp. 174–177.
[39] R. Huerta, S. Vembu, J.M. Amigó, T. Nowotny, C. Elkan, Inhibition in multiclass classification, Neural Comput. 24 (9) (2012) 2473–2507.
[40] Figaro USA, http://www.figarosensor.com/.
[41] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed. Wiley-Interscience, 2000.
[42] Bronkhorst High-Tech B.V., http://www.bronkhorst.com/.
[43] Airgas Inc., http://www.airgas.com/.
[44] The LabVIEW Environment, http://www.ni.com/labview/.
[45] A. Vergara, M.K. Muezzinoglu, N. Rulkov, R. Huerta, Information-theoretic optimization of chemical sensors, Sens. Actuators, B 148 (1) (2010) 298–306.
[46] N. Barsan, U. Weimar, Conduction model of metal oxide gas sensors, J. Electroceram. 7 (3) (2001) 143–167.
[47] G. Korotcenkov, Metal oxides for solid-state gas sensors: what determines our choice? Mater. Sci. Eng. B 139 (1) (2007) 1–23.
[48] M. Muezzinoglu, A. Vergara, R. Huerta, N. Rulkov, M. Rabinovich, A. Selverston, H. Abarbanel, Acceleration of chemo-sensory information processing using transient features, Sens. Actuators, B 137 (2) (2009) 507–512.
[49] R. Lomasky, Active Acquisition of Informative Training Data, Tufts University, 2009, (Ph.D. thesis).
[50] R. Lomasky, C. Brodley, M. Aernecke, D. Walt, M. Friedl, Active class selection, ECML, Vol. 4701 of Lecture Notes in Computer Science, Springer, 2007, pp. 640–647.
[51] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[52] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, 1992, pp. 144–152.
[53] S. Dasgupta, Two faces of active learning, Theor. Comput. Sci. 412 (19) (2011) 1767–1781.
[54] P. Clifford, D. Tuma, Characteristics of semiconductor gas sensors I. Steady state gas response, Sensors Actuators 3 (1983) 233–254.
[55] P. Clifford, D. Tuma, Characteristics of semiconductor gas sensors II. Transient response to temperature change, Sensors Actuators 3 (1983) 255–281.
[56] J.D. Ingle Jr., S.R. Crouch, Spectrochemical Analysis, Old Tappan, Prentice Hall College Book Division, NJ (US), 1988.
[57] H. Wohltjen, R. Dessy, Surface acoustic wave probe for chemical analysis. I. Introduction and instrument description, Anal. Chem. 51 (9) (1979) 1458–1464.
[58] J.R. Stetter, J. Li, Amperometric gas sensors a review, Chem. Rev. 108 (2) (2008) 352–366.
[59] J. Firth, A. Jones, T. Jones, The principles of the detection of flammable atmospheres by catalytic devices, Combust. Flame 20 (3) (1973) 303–311.
[60] J.S. Murguía, A. Vergara, C. Vargas-Olmos, T.J. Wong, J. Fonollosa, R. Huerta, Two-dimensional wavelet transform feature extraction for porous silicon chemical sensors, Anal. Chim. Acta. 785 (2013) 1–15.
[61] J. Janata, Principles of Chemical Sensors, Springer, 2009.
[62] E.L. Allwein, R.E. Schapire, Y. Singer, Reducing multiclass to binary: a unifying approach for margin classifiers, J. Mach. Learn. Res. 1 (2001) 113–141.
[63] P.L. Bartlett, M. Traskin, AdaBoost and other large margin classifiers: convexity in pattern classification, Proceedings of the 5th Workshop on Defence Applications of Signal Processing, 2006.
[64] S. Canu, A. Smola, Kernel methods and the exponential family, Neurocomputing 69 (7) (2006) 714–720.
[65] P. Pletscher, C.S. Ong, J.M. Buhmann, Entropy and margin maximization for structured output learning, Machine Learning and Knowledge Discovery in Databases, Springer, 2010, pp. 83–98.
[66] I. Rodriguez-Lujan, C. Santa Cruz, R. Huerta, On the equivalence of kernel Fisher discriminant analysis and kernel quadratic programming feature selection, Pattern Recogn. Lett. 32 (11) (2011) 1567–1571.
[67] P.L. Bartlett, M.I. Jordan, J.D. McAuliffe, Convexity, classification, and risk bounds, J. Am. Stat. Assoc. 101 (473) (2006) 138–156.
[68] A. Tewari, P.L. Bartlett, On the consistency of multiclass classification methods, J. Mach. Learn. Res. 8 (2007) 1007–1025.
[69] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, neural networks, IEEE Trans. 12 (2) (2001) 181–201.