Sensors and Actuators B 185 (2013) 201–210

Contents lists available at SciVerse ScienceDirect

Sensors and Actuators B: Chemical

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

Active temperature modulation of metal-oxide sensors for quantitative analysis of gas mixtures

Rakesh Gosangi [a,*], Ricardo Gutierrez-Osuna [b,1]

[a] Department of Computer Science and Engineering, Texas A&M University, TAMU 3112, College Station, TX 77843-3112, United States
[b] Department of Computer Science and Engineering, Texas A&M University, 506A H.R. Bright Building, College Station, TX 77843-3112, United States

ARTICLE INFO

Article history:
Received 28 March 2012
Received in revised form 24 March 2013
Accepted 15 April 2013
Available online 28 April 2013

Keywords:
Active sensing
Metal-oxide sensors
Multicomponent analysis
Gaussian mixture models

ABSTRACT

We present an active sensing method for quantitative analysis of gas mixtures using metal-oxide (MOX) chemical sensors. The method allows a MOX sensor to adapt its operating temperature in real time so as to sequentially reduce uncertainty in the concentration estimates. We formulate the problem as one of probabilistic state estimation coupled with a myopic optimization algorithm. At each iteration, the algorithm estimates the expected reduction in entropy for each sensing action (i.e., operating temperature) and selects the best such temperature. We first evaluated the proposed method on a simulated binary mixture problem using a computational model of MOX sensors. In these simulations, we compared the active sensing approach against conventional sequential forward selection (SFS) strategies. We then experimentally validated the method on a Taguchi gas sensor to quantify mixtures of two organic compounds. Our results indicate that the active sensing algorithm can obtain comparable estimation performance as SFS with significantly fewer measurements. In addition, since active sensing selects features on the fly, it is also more robust to experimental noise than off-line subset selection strategies.

1. Introduction

Metal-oxide (MOX) gas sensors are robust, inexpensive, and highly sensitive but have poor selectivity [1]. As a result, discriminating among several chemicals, or between targets and backgrounds, becomes difficult. The issue can be addressed in part by combining multiple sensors with different selectivities, though the sensor-array response tends to be highly collinear, which leads to numerical problems. An alternative approach has been used by a number of researchers, and consists of modulating the sensor's operating temperature while it is exposed to a gas sample. Since the selectivity of MOX materials is dependent on the operating temperature, the sensor responses thus obtained have gas-specific signatures that can be used to improve the sensor's selectivity [2–4]. Temperature modulation has also been used in quantitative analysis of gas mixtures (also referred to as multicomponent analysis); early work by [5] shows that the resulting responses are characteristic of the components present in the mixture.

In this paper, we explore the problem of optimizing temperature programs for multicomponent analysis. Namely, given a mixture of known non-interacting gases, we seek to find a temperature sequence for a MOX sensor that best estimates the concentration of the individual components. The approach builds on our prior work on active sensing for classification problems [6], where we used Markov decision processes to optimize temperature programs in real time. Here, we extend the idea of active sensing for quantitative analysis of gas mixtures.

We formulate the problem as that of probabilistic state estimation, where each state represents a concentration profile. We maintain a belief distribution that assigns a probability to each profile, and update the belief distribution at each time step by incorporating the latest sensor measurement. To select the sensor's next operating temperature, we use a myopic algorithm based on information-theoretic measures. Namely, the myopic algorithm selects the operating temperature that is best expected to reduce the uncertainty in the future belief distribution. Preliminary results of this work on simulated data were presented at ISOEN 2011 as an extended two-page abstract [7]. Here, we present the complete mathematical formulation, benchmark the approach against feature subset selection strategies, and validate it experimentally on a commercial MOX sensor.

The rest of the paper is organized as follows. Section 2 provides a literature review on quantitative analysis of gas mixtures with MOX sensors and a brief background on active sensing. In Section 3, we present the mathematical formulation for the problem and describe the proposed active-sensing algorithm. In Section 4, we evaluate the method on a simulated binary mixture problem using a computational model of MOX sensors, and compare it

* Corresponding author. Tel.: +1 551 200 8907.
E-mail addresses: rakesh@cse.tamu.edu, rakesh.gosangi@gmail.com (R. Gosangi), rgutier@cse.tamu.edu (R. Gutierrez-Osuna).
[1] Tel.: +1 979 845 2942; fax: +1 979 847 8578.

0925-4005/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.snb.2013.04.056

against conventional feature subset selection strategies. Section 5 describes the experimental setup used for validating the approach on commercial MOX sensors and presents experimental results. We conclude the paper with a discussion and directions for future work.

2. Background

The problem of quantitative analysis of gas mixtures with MOX sensors has been studied for more than two decades [8,9]. The typical approach is to build a mapping from the independent variables (sensor responses) to the dependent variables (concentrations). Both linear (e.g. ordinary least squares) and nonlinear (e.g. neural networks) have been used for this purpose. The conventional approach for building linear models is based on linear regression. In linear regression, concentrations are modeled as a weighted linear combination of the sensor responses, and the weights are estimated in a least-squares fashion by minimizing the sum-squared prediction error over training data. These weights could be further stabilized using methods such as principal components regression [10,11], ridge regression [12], or partial least squares [13–15]. Nonlinear regression models have also been used based on several artificial neural networks architectures, including multi-layer perceptrons [9,16], self-organizing maps [17], and time delay neural networks [18,19].

In one of the earliest studies on temperature modulation for multicomponent analysis, Nakata et al. [5] applied sinusoidal temperature programs to SnO_2 sensors in order to quantify mixtures of hydrocarbons. The authors analyzed changes in sensor conductance with Fast Fourier Transform (FFT). Their results showed that the amplitudes of the FFT harmonics exhibited characteristic changes with analyte concentration, and could be used for quantifying the hydrocarbons in mixtures. Heilig et al. [20] used a similar approach to analyze binary mixtures of CO and NO_2 in air. Llobet et al. [11] conducted similar studies with SnO_2 sensors for quantifying mixtures of CO and NO_2. In contrast with previous studies, the authors used discrete wavelet transforms (DWT) to extract features from the sensor responses. Their results suggest that DWT outperforms FFT in terms of estimation accuracy. More recently, Vergara et al. [21] presented a systematic approach to optimize temperature programs applied to microhotplate metal-oxide gas sensors for multicomponent analysis. The authors used a multisinusoidal signal for temperature modulation and trained a PLS regression model to predict the constituent concentrations from sensor responses. The authors then selected the best operating frequencies using 5-fold cross validation to quantify different mixtures of ethylene, ammonia, and acetaldehyde.

2.1. Active sensing

The idea of active sensing (Fig. 1) can be traced back to the robotics literature, where it was defined as 'controlling strategies applied to the data acquisition process which will depend on the current state of the data interpretation and the goal or the task of the process' [22]. In robotics, active sensing is commonly used in localization and navigation tasks, where there is a trade-off between the immediate rewards of actions (e.g. bringing the robot closer to its goal) and long-term effects (e.g. gathering information to avoid getting lost along the way) [23–26]. Among other robotics applications, active sensing has also been used for simultaneous localization and mapping [27], off-road driving [28], and robotic exploration [29]. In the last decade, most work in active sensing has been centered on military applications such as target tracking using remote radar, sonar, and electro-optical sensors [30]. These methods have also been applied to problems such as landmine detection [31], underwater mine classification [32], and dynamic target tracking with wireless sensor networks [33,34].

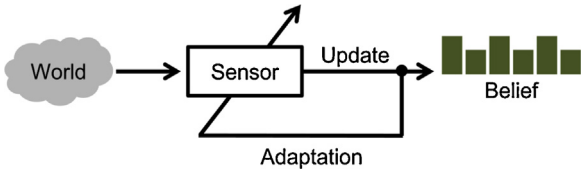

Fig. 1. Inactive sensing the system adapts its sensing parameters based on its belief about the state of world, which in turn is updated based on information obtained from the sensors.

A handful of research groups have also explored active-sensing strategies in the domain of chemical sensors. In one of the earliest studies, Nakamoto et al. [35] developed a method for active odor blending, where the goal was to reproduce an odor blend by creating a mixture from its individual components. The authors developed a control algorithm that adjusted the mixture ratio so that the response of a gas sensor array to the mixture matched the response to the odor blend. Priebe et al. [36] developed a statistical pattern recognition method based on the concept of integrated sensing and processing (ISP). Given a feature vector, the method builds a decision tree that partitions feature space hierarchically; nodes close to the root of the tree select features based on their ability to provide good clustering of examples regardless of class labels, whereas nodes at the leaves select features based on their ability to discriminate examples from different classes. The authors evaluated the model on an experimental dataset from an array of 19 optical sensors exposed to trichloroethylene (a carcinogenic industrial solvent) in complex backgrounds. Their results show that the ISP method can reduce misclassification rates by 50%, while requiring only 20% of all the sensors to make any individual classification. More recently, Dinakarababu et al. [37] proposed an adaptive spectroscopic architecture called adaptive feature specific spectrometer (AFSS). Unlike a traditional IR spectrometer, AFSS has a tunable spectral filter (a digital micromirror device), which aids in multiplexing certain spectral bands and directing them onto a photo-detector. The system thus measures the projection of the incoming spectral density onto a set of basis vectors, rather than measuring the spectral density directly. The basis vectors are changed over time based on information from previous measurements. Additional work on adaptive chemosensing systems can also be found in a recent review article [38].

3. Methods

Given a mixture of n known non-interacting gases and a MOX sensor that operates at D different temperatures $\boldsymbol{A} = (a_1, a_2, \ldots, a_D)$, we seek to estimate the concentrations $\theta = (c_1, c_2, \ldots, c_n)$ of the n gases by applying a sequence of operating temperatures $(a_1, a_2, \ldots, a_t)$. For simplicity, we assume that each gas concentration c_i can take one of k_i discrete values.[2] As a result, the state of the system can be represented by a vector θ with $r = k_1 \times k_2 \times \cdots \times k_n$ different admissible values: $\theta \in S = \{s_1, s_2, \ldots, s_r\}$, where s_i is an n-dimensional vector containing the discrete concentrations of the n gases.

We solve this problem through probabilistic state estimation by maintaining a belief distribution over the r possible concentration

[2] This assumption allows us to constrict the representation to a discrete state-space, which greatly simplifies the optimization problem from a computational standpoint.

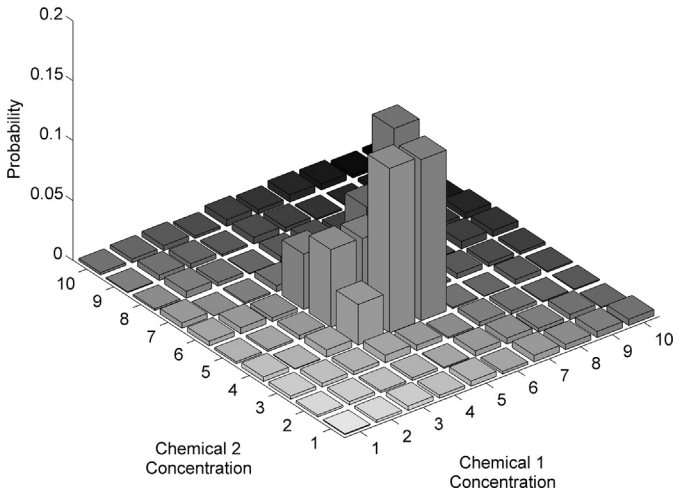

Fig. 2. Example of a belief distribution over the 'concentration space' for a binary mixture.

profiles. The belief b_t at time t assigns a probability to each one of the r states:

$$b_t : \boldsymbol{S} \rightarrow [0, 1]; \quad \sum_{s_i \in \boldsymbol{S}} b_t(s_i) = 1 \tag{1}$$

Fig. 2 shows an example of a hypothetical belief distribution over 100 (10×10) profiles for a mixture of two chemicals. Given an initial belief $b_0(s)$, a sequence of operating temperatures $\boldsymbol{a} = (a_1, a_2, \ldots, a_t)$ and the corresponding observations $\boldsymbol{o} = (o_1, o_2, \ldots, o_t)$[3], the current belief $b_t(s)$ is defined as:

$$b_t(s) = p(s|o_1, o_2, \ldots, o_t, a_1, a_2, \ldots, a_t) \tag{2}$$

We assume that the belief $b_t(s)$ acts as a *sufficient statistic*, and can be updated using the previous estimate $b_{t-1}(s)$:

$$b_t(s) = p(s|o_t, a_t, b_{t-1}) \tag{3}$$

Our goal is to select a sequence of operating temperatures that minimizes uncertainty in θ, which can be estimated as:

$$H(\theta) = -\sum_{s_i \in \boldsymbol{S}} p(\theta = s_i) \log(p(\theta = s_i)) \tag{4}$$

where $H(\theta)$ is the Shannon entropy. In what follows, we will use the notation $H(\theta_t)$ to represent the entropy of θ at time t:

$$H(\theta_t) = -\sum_{s_i \in \boldsymbol{S}} p(\theta_t = s_i) \log(p(\theta_t = s_i)) = -\sum_{s_i \in \boldsymbol{S}} b_t(s_i) \log(b_t(s_i)) \tag{5}$$

An exact solution to this problem (called a policy π) is a mapping from a belief distribution b_t to an operating temperature $\pi: b_t \rightarrow \boldsymbol{A}$. However, finding an exact solution is P-SPACE complete [39] and thus computationally expensive. Also, an exact solution would allow for repeating operating temperatures which is not desired in the current scenario. Therefore, we use a myopic approach that optimizes the temperature program on a per-temperature basis. Our myopic algorithm selects the next operating temperature a_{sel} as the one with the largest expected reduction in entropy:

$$a_{sel} = \underset{a_{t+1}}{arg\,\max}(H(\theta_t) - H(\theta_{t+1}|a_{t+1})) \tag{6}$$

[3] o_t does not necessarily have to be the steady-state response of the sensor. It could be any feature (or features) extracted from the transient response of the sensor to action a_t.

Table 1
Flowchart of the algorithm used for optimizing temperature programs.

- Step 1: Initialize a uniform belief distribution $b_0(s)$, and a set of available configurations $\boldsymbol{A}$
- Step 2: Select the "best" operating temperature $a_{sel} = \text{argmax}_{a_{t+1} \in \boldsymbol{A}}(H(\theta_t) - H(\theta_{t+1}|a_{t+1}))$; $t = t + 1$
- Step 3: Operate the sensor at $a_t = a_{sel}$ and measure the sensor response o_t, remove a_{sel} from $\boldsymbol{A}$
- Step 4: Update the belief distribution $b_t(s) = p(s|o_t, a_t, b_{t-1})$
- Step 5: Measure the entropy $H(\theta_t)$
  - If $H(\theta_t) \geq \psi$ go to Step 2
  - Else declare $\theta = \text{argmax}_s\,(b_t(s))$

where $H(\theta_{t+1}|a_{t+1})$ is the expected entropy of θ following application of temperature a_{t+1}. After selecting the operating temperature, the system transitions to the next time step (i.e., $t = t + 1$). Then, we operate the sensor at temperature $a_t = a_{sel}$ and measure the resulting response o_t; this observation o_t is used to update the belief distribution $b_t(s) = p(s|o_t, a_t, b_{t-1})$. In a final step, we estimate the entropy of the updated belief distribution. If the entropy is below a pre-determined threshold ψ, the concentration is estimated as $\theta = \text{argmax}_s b_t(s)$; this halts the sensing process and avoids any further data acquisition. If on the other hand the entropy exceeds the threshold ψ, a new operating temperature is selected, and the process is repeated. A pseudo-code of the myopic algorithm is shown in Table 1. This pseudo-code leaves two questions unanswered: (1) How do we update the belief distribution? (2) How do we estimate the expected entropy $H(\theta_{t+1}|a_{t+1})$?

3.1. Updating the belief distribution

Given belief b_{t-1}, and action a_t leading to observation o_t, we obtain the updated belief distribution b_t using a recursive Bayesian filter [40]:

$$b_t(s) = \frac{p(o_t|a_t, s)\sum_{s' \in \boldsymbol{S}} b_{t-1}(s')p(s|s', a_t)}{p(o_t|a_t, b_{t-1})} \tag{7}$$

where $p(o_t|a_t, s)$ is obtained from a probabilistic model of the sensor (discussed in Section 3.1.1), $p(s|s', a_t)$ is the probability that θ changes from s to s' upon taking action a_t and $p(o_t|a_t, b_{t-1})$ acts as a normalization term which ensures the belief distribution sums to 1. Since the concentration profile θ does not change over time, the transition model becomes:

$$p(s|s') = \begin{cases} 1 & \text{if } s = s' \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

As a result, the belief update simplifies to:

$$b_t(s) = \frac{p(o_t|a_t, s)b_{t-1}(s)}{p(0_t|a_t, b_{t-1})} \tag{9}$$

At time $t = 0$, when no sensor measurements are obtained, we assume that all the concentration profiles are equally likely, that is, $b_0(s) = 1/r, \quad \forall s \in \boldsymbol{S}$.

3.1.1. Building the sensor model

The term $p(o_t|a_t, s)$ denotes the probability of observing sensor response o_t at concentration setting s when the sensor is operated at temperature a_t. To enable real-time estimation of concentrations, it is crucial that this likelihood value be computed efficiently. For this purpose, we model the response of the sensor o_{ij} at a temperature a_i to concentration profile s_j with a Gaussian mixture model (GMM) as follows:

$$p(o_{ij}|a_i, s_j) = \sum_{m=1}^{M} \alpha_m^{(i,j)} N(o_{ij}|\mu_m^{(i,j)}, \sigma_m^{(i,j)}) \tag{10}$$

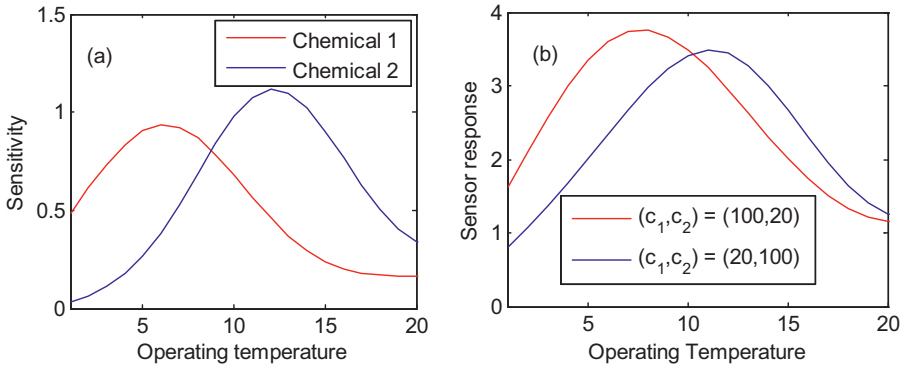

Fig. 3. (a) Temperature-dependent sensitivity of the simulated MOX sensor to two hypothetical chemicals. (b) Response to two different binary mixtures when the sensor temperature was ramped from 1 to 20.

where M is the total number of Gaussians, and $\alpha_m^{(i,j)}$, $\mu_m^{(i,j)}$, $\sigma_m^{(i,j)}$ are the mixing coefficient, mean, and covariance matrix for each Gaussian, respectively. Here, state s_j and temperature a_i are discrete variables but the sensor response o_{ij} is a continuous variable. GMMs are suited for this purpose since they are universal approximators for continuous distributions: given a sufficient number of mixture components, GMMs can model any continuous distribution [41].

Given training data, we use expectation maximization [42] to learn model parameters that maximize the likelihood of the data. The training data for each mixture is generated by running the sensor with a set of random temperature sequences and recording the corresponding responses. Each random sequence contains all the operating temperatures, with each temperature occurring once, and the sequences are different for each mixture. During the testing phase, the sensor will be driven with a temperature sequence that is most likely not part of the training set. Thus, it is important that the training set captures as much of the variance introduced due to the modulation process as possible, and for this reason we use randomized rather than fixed sequences.

3.2. Estimating the expected entropy

The term $H(\theta_{t+1}|a_{t+1})$ denotes the expected entropy in the state variable θ if the sensor were to be operated at temperature a_{t+1}. We estimate the expected entropy as the average over all possible sensor observations weighted by the probability of each observation:

$$H(\theta_{t+1}|a_{t+1}) = \sum_{\forall o_{t+1}} p(o_{t+1}|a_{t+1}) H(\theta_{t+1}|a_{t+1}, o_{t+1}) \tag{11}$$

where $p(o_{t+1}|a_{t+1})$ is the probability of making observation o_{t+1} at temperature a_{t+1} irrespective of concentration, and $H(\theta_{t+1}|a_{t+1}, o_{t+1})$ is the entropy of θ at time $t+1$ following observation o_t+1 at temperature a_{t+1}. Estimating $H(\theta_{t+1}|a_{t+1})$ can be interpreted as projecting uncertainty into the immediate future, and involves the following three steps. First, for each possible observation o_t+1, we use the Bayesian update in Eq. (9) to calculate the projected belief distribution b_t+1 using the current belief b_t and action a_{t+1}. Then, for each projected belief b_t+1, we compute the expected Shannon entropy in Eq. (5); this produces $H(\theta_{t+1}|a_{t+1}, o_{t+1})$. Finally, we sum all these entropies, weighted by the corresponding observation probabilities $p(o_{t+1}|a_{t+1})$, to estimate the expected entropy $H(\theta_{t+1}|a_{t+1})$ for action a_{t+1}.

To estimate $p(o_{t+1}|a_{t+1})$, we create concentration-independent sensor models during the training stage. Namely, we collect all the training observations (irrespective of the sample concentrations) corresponding to each operating temperature and model them with a GMM. These GMMs have $M \times |\boldsymbol{S}|$ components, where M is the number of components used for the concentration-dependent GMMs (as in Section 3.1.1), and $|\boldsymbol{S}|$ is the number of admissible values for θ Using these models, $p(o_{t+1}|a_{t+1})$ can be estimated as in Eq. (10).

The expected entropy calculation in Eq. (11) is only applicable for discrete observations; for continuous spaces, the summation in Eq. (11) becomes an indefinite integral. To address this issue, we discretize the observation space for each temperature using k-means clustering.[4] This results in a finite set of (sorted) cluster centers $\{o_{a_i,1}, o_{a_i,2}, \ldots, o_{a_i,k}\}$ for each temperature a_i. From here, we estimate the likelihood values $p(o_{a_i,j}|a_i)$ using the sensor model described in Section 3.1.1, and the likelihood values are then used in Eq. (11) to calculate the expected entropy.

3.3. Incorporating sensing costs

The above formulation does not take sensing costs into consideration. However, each sensing action (temperature a_{t+1}) could have an associated sensing cost $C(a_{t+1}) = c_{t+1}$ to reflect the amount of power consumed at that temperature, the time required for the temperature pulse to reach steady state, etc. These sensing costs can be incorporated into the formulation by modifying the temperature selection criterion of Eq. (6) as follows:

$$a_{sel} = \underset{a_{t+1}}{arg\,max} U(a_{t+1}) = \underset{a_{t+1}}{arg\,max}((H(\theta_t) - H(\theta_{t+1}|a_{t+1}) - \alpha \times c_{t+1}) \tag{12}$$

where $U(a_{t+1})$ denotes the utility of temperature a_{t+1}, defined as a weighted sum of the sensing cost c_{t+1} and the expected reduction in entropy; the parameter α provides a trade-off between the two criteria (entropy and sensing cost) and serves as a normalization constant to ensure that they have comparable magnitudes. If the utility of all temperatures is negative, there is no incentive in any further data acquisition. This happens when the sensing costs of all temperatures exceed their expected reductions in entropy. Under this condition, we halt the sensing process and declare the concentration profile as $\theta = \text{argmax}_s(b_t(s))$. For the experimental work presented in the rest of the paper, we assumed $\alpha = 0$.

4. Validation on synthetic data

We first validated the proposed active sensing (AS) method on a simulated MOX sensor model through a series of comparisons with SFS [43]. These experiments allowed us test the

[4] We chose k-means clustering for the discretization process because it provides a more accurate representation of the distribution in feature space than uniform discretization and is also robust to outliers.

method extensively; such comprehensive evaluation is impractical on experimental data due to the prolonged time required for data collection. In contrast to AS where features are selected online, SFS is a passive strategy that selects an near-optimal feature subset off-line using training data. SFS operates in a greedy fashion: it starts with an empty feature set and sequentially adds the feature that, when combined with all features selected previously, maximizes an objective function. In our case, the objective function was the classification performance of a naïve Bayes classifier with 10-fold cross validation; this is known as a "wrapper" in the feature-selection literature [44] and prevents SFS from over-fitting the training data.

4.1. Simulated sensor model

We used a simulated sensor model to generate a dataset consisting of sensor responses to mixtures of two non-interacting chemicals. Following Clifford and Tuma [45], we modeled the sensor response as:

$$o_t = \sum_i S_i(a_t) c_i^{\beta_i} + \gamma o_{t-1} \tag{13}$$

where a_t denotes temperature at time t, o_t is the sensor response, $S_i(a_t)$ is the sensor sensitivity to gas i at temperature a_t, c_i is the concentration of gas i, β_i is a gas-dependent parameter, and γ is a parameter that captures sensor dynamics (history effects). Following [46], we defined the sensitivity S_i of the sensor to a gas i as a sum of a univariate Gaussian and a linear function over the operating temperatures:

$$S_i(a_t) = l_i a_t + \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp -\frac{(a_t - \tau_i)^2}{2\sigma_i^2} \tag{14}$$

where l_i, σ_i, and τ_i are gas-dependent parameters.

We created an experimental scenario with 20 different temperatures $\{1, 2, \ldots, 20\}$ and two hypothetical chemicals: $\{l_1, \sigma_1, \tau_1\} = \{0.1, 0.4, 12\}$ and $\{l_2, \sigma_2, \tau_2\} = \{0.08, 0.45, 6\}$; parameters in the Clifford–Tuma model were set to $(\beta_1, \beta_2, \gamma) = (0.45, 0.5, 0.1)$. Fig. 3(a) shows the temperature-sensitivity profile of the simulated MOX sensor for two hypothetical chemicals. We assumed each chemical could be present at one out of five discrete concentrations ($c_1, c_2 \in \{20, 40, 60, 80, 100\}$), resulting in $r = 5 \times 5 = 25$ different concentration profiles.

4.2. Experiment I: performance vs. number of actions

In a first experiment, we compared the two methods (AS and SFS) as a function of the number of operating temperatures used. For this purpose, we created five training datasets each containing 125 samples: for each of the 25 concentration profiles, we generated five random permutations of the 20 operating temperatures and then obtained sensor responses with Eq. (13); o_0 was initialized to a random value. From these datasets, we trained five sets of GMM sensor models described in Section 3.1.1, and selected SFS feature subsets of different cardinalities f ($1 \le f \le 10$). At each cardinality f and for each set of sensor models, we tested both methods on a dataset of 250 samples (10 per concentration profile). To ensure a fair comparison between both methods, we modified AS such that the algorithm stopped sensing when either f observations were utilized or $H(\theta_t) < 0.65$[5]; in addition, we did not allow AS to use the same feature twice.

[5] We chose the value of 0.65 for ψ because it reflects a situation with 90% certainty in the classification. In other words, consider a belief distribution that assigns a probability of 0.90 to one of the concentration profiles and $0.004 = 0.1/24$ to each of the 24 remaining profiles. This distribution has entropy of 0.65 nat.

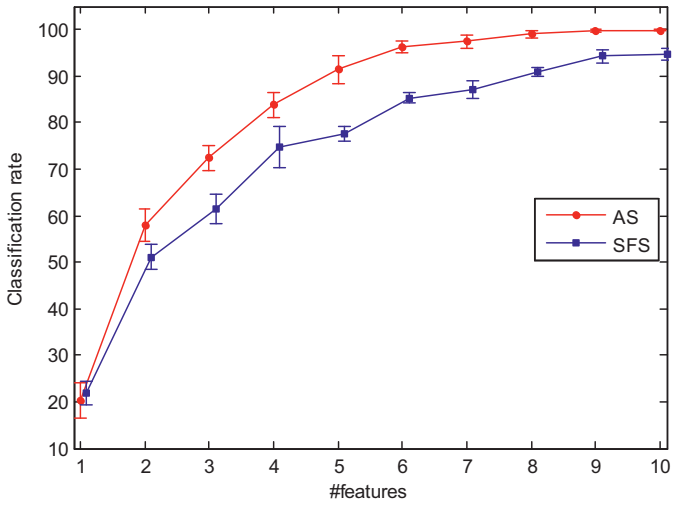

Fig. 4. Classification performance of both methods as a function of the number of features used. The SFS curve has been shifted right along the x-axis to avoid overlapping of error bars.

Results are summarized in Fig. 4. As expected, classification performance for both methods improved with increasing number of features f. Interestingly, when only one temperature is used ($f = 1$), SFS outperformed AS: 21.9% vs. 20.3%, respectively. This is likely due to the fact that SFS uses 10-fold cross validation to identify the initial temperature, where as AS selects the feature based solely on the sensor models and has no provision to cross-validate the selection. However, for $f > 1$ AS consistently outperforms SFS, and achieves perfect classification at $f = 7$. The superior performance of AS may be attributed to its adaptive nature, which allows it to change the temperature program based on information obtained thus far from the test sample, whereas SFS always uses a pre-determined set of operating temperatures that were optimized off-line.

4.3. Experiment II: performance vs. noise

In a second experiment, we compared the two methods as a function of measurement noise. As before, we generated five training datasets (each containing 125 samples) using the simulated sensor model, and then trained the GMM sensor models. We then

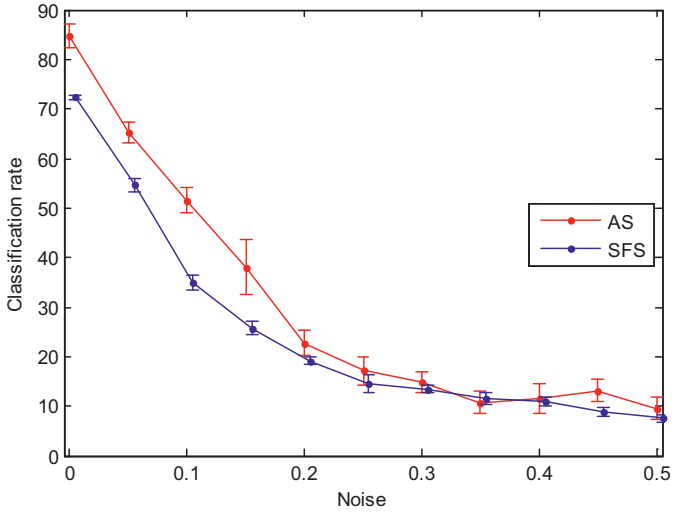

Fig. 5. Classification performance of both methods versus variance of additive Gaussian noise. The SFS curve has been shifted right along the x-axis to avoid overlapping of error bars.

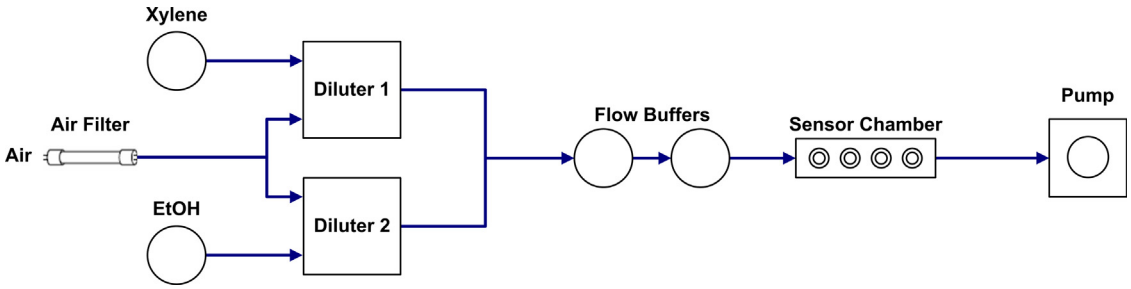

Fig. 6. A schematic of the delivery system used for experimentation.

used SFS to identify the best feature subsets of cardinality $f = 5$. Following the previous experiment, the AS algorithm stopped sensing when either 5 observations were obtained or $H(\theta_t) < 0.65$; AS was also prevented from using the same feature twice. To test the two methods, we generated 11 different test sets, each having additive Gaussian noise with variance ranging from $0.0 \leq \sigma^2 \leq 0.5$ in steps of 0.05. Each test dataset contained 250 test samples, 10 per concentration profile.

Classification results are summarized in Fig. 5. As expected, classification performance for both methods degrades with increasing levels of noise in the test data. However, in the noise range $0.0 \leq \sigma^2 \leq 0.3$, AS consistently outperforms SFS. This is because AS selects features at measurement time, which allows it to adapt to noise, whereas SFS uses a pre-specified set of temperatures that was computed off-line using noise-free training data. The average number of features used by AS increases with decreasing SNR because; the increasing levels of noise do not allow the beliefs to reach the certainty levels required for AS to stop the sensing process. Therefore, AS acquires more features to 'counter' the noise, whereas SFS does not have this capability. In addition, with increasing levels of noise, the observations are more likely to fall in areas where the class distributions overlap. In such situations, the beliefs associated with these overlapping classes would have similar values. Then, for the subsequent sensing steps, AS will give preference to features that can strongly distinguish between the 'overlapping classes'. In contrast, SFS works with a predefined sequence of features. At higher levels of measurement noise ($0.35 \leq \sigma^2 \leq 0.5$) both methods perform comparably, indicating that at these settings the amount of noise in the data prevents AS from making accurate predictions.

5. Validation on experimental data

5.1. Experimental setup

We also validated the AS algorithm experimentally using a commercial MOX sensor (TGS 2620[6]) with six binary mixtures of xylene and ethyl-alcohol (EtOH) introduced into a background of desiccated air. The sensor was placed in a custom-built sensor chamber. The sensor's measuring circuit was a voltage divider, where the sensor was connected in series with a load resistor ($R_L = 5\,k\Omega$) and the input circuit voltage was $V_c = 10\,V$. Gas mixtures were delivered from 30 ml glass vials using an air pump connected downstream. The concentrations of the two chemicals were controlled using two Model 1010 Precision Gas diluters.[7] A schematic of the experimental setup is shown in Fig. 6. Sensor and diluters were interfaced using NI-6024E and NI-6713[8] data acquisition cards and controlled with Matlab.

[6] http://www.figarosensor.com/.
[7] http://www.customsensorsolutions.com/.
[8] http://www.ni.com/dataacquisition/pci/.

Table 2
Concentrations of the two chemicals in the mixtures, expressed in volume/volume.

Mixture	Xylene concentration (%)	EtOH concentration (%)
1	5	5
2	10	5
3	15	5
4	5	15
5	10	15
6	15	15

We conducted a preliminary study (details not included in this paper) to determine a suitable range for the concentration of the chemicals and sensor heater voltage; our goal was to ensure the quantification problem was not trivial so that temperature modulation would be needed to discriminate the mixtures. The resulting binary mixtures are summarized in Table 2.

The sensor was operated at 10 different heater voltages ranging from $4.5\,V \leq V_H \leq 7.5\,V$ in steps of 0.33 V, which corresponds to a temperature scale of 320–550 °C. At each operating voltage, we pulsed the sensor for a duration of 30 s. Between consecutive pulses, the sensor was reset to a baseline voltage (0 V), also for duration of 30 s. This form of temperature programming helped reduce variance in the sensor response due to thermal dynamics. Fig. 7 shows a sample sensor response to a mixture of 5% xylene and 5% EtOH for a random sequence of 5 V pulses. For each binary mixture, we operated the sensor with 12 random sequences, each sequence containing the 10 operating voltages; this resulted in a dataset with 72 samples (6 mixtures × 12 sequences). The mixtures were presented in a randomized order to avoid systematic errors.

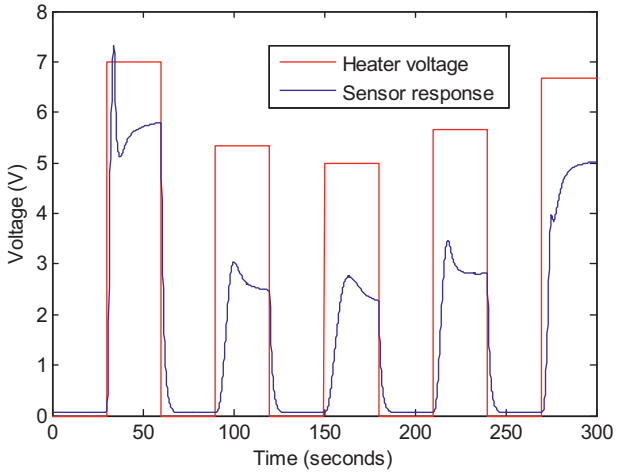

Fig. 7. Transient sensor response to a mixture of 5% xylene and 5% EtOH. The sensor heater was driven with a sequence of 5 V pulses, each 30 s long feature extraction.

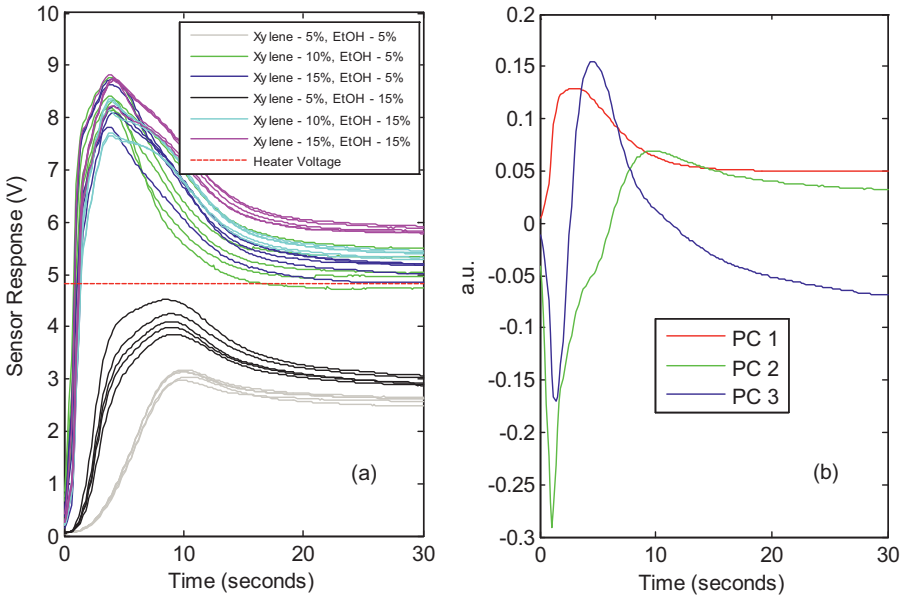

Fig. 8. (a) Sensor transient to the six gas mixtures (5 repetitions per mixture) in response to a 4.83 V pulse in heater voltage. This pulse was preceded and succeeded by a 0 V pulse of 30-s duration. (b) The first three principal components extracted from the transients in (a).

5.2. Feature extraction

We used principal component analysis (PCA) [47] to extract features from the sensor transients. Namely, for each of the 10 V settings ($4.5\,V \le V_H \le 7.5\,V$; $\Delta = 0.33\,V$), we collected the sensor responses to all binary mixtures and then applied PCA to obtain the loadings (eigenvectors) and scores. Fig. 8(a) shows 30 sensor transient responses to the six binary mixtures (5 transients per mixture) to a 30-s 4.83 V pulse; Fig. 8(b) shows the first three loadings. We repeat this process for each voltage setting to obtain the corresponding loadings and scores. In all cases, the first three principal components were sufficient to capture more than 99% of the variance. Thus, PCA allowed us to compress the transient response down to three features (i.e., the PCA scores). Then, we used three 1D GMMs (one per principal component) to create the probabilistic sensor models[9]; refer to Section 3.1.1 for details. During the testing stage, when the sensor is driven with a pulse a_t, the resulting transient response is multiplied with the PCA loadings corresponding to a_t. The scores thus obtained are treated as the observation o_t.

5.3. Experiment III: AS vs. SFS

Following procedures described in Section 4, we compared the AS algorithm against SFS. To avoid over-fitting, we used a naïve Bayes classifier wrapped in a10-fold cross-validation loop as the objective function for SFS. For each fold, we randomly divided the sensor data (72 samples) into two subsets: a training dataset containing 30 samples (5 per mixture), and a test dataset containing 42 samples (7 per mixture). Then, we ran SFS on the training datasets to generate optimal feature subsets of cardinalities $2 \le f \le 7$. For each of these subsets, we estimated classification performance on the corresponding test datasets. We used the same 10-fold cross validation loop to estimate the performance of AS. Namely, we trained GMM sensor models on each of the 10 training sets and estimated the performance of AS on the corresponding test sets. To ensure a fair comparison between the two methods, the AS algorithm was stopped when either f observations were obtained or $H(\theta_t) < 0.1$[10]; as earlier, AS was not allowed to choose the same feature twice.

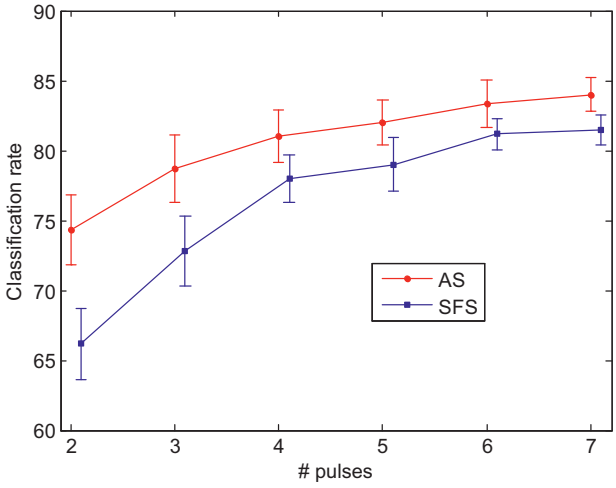

Fig. 9. Classification performance of AS and SFS as a function of number of features used. For visualization purposes, the SFS curve was shifted slightly along the x-axis.

Results are summarized in Fig. 9. As expected, classification performance for both methods improved with increasing number of features. However, AS consistently outperformed SFS regardless of feature set size. The disparity between the two methods is more evident at lower cardinalities; as the number of features increases, there is a greater overlap between the cumulative feature sets selected by the two methods. These results are consistent with those obtained on simulated data (see Fig. 4).

We calculated the average number of pulses used by AS for all the mixtures (across all rounds of cross validation and for $f = 7$). The results are summarized in Table 3. For mixtures #1 and #4, AS requires an average of 2 pulses to identify the samples. Interestingly, these are the two mixtures with 5% Xylene. For mixture #6 (15% both chemicals), AS requires an average of 3 pulses to identify

[9] We chose to use three 1D GMMs as opposed to using 3D GMMs due to the limited number of training samples available.

[10] The value of 0.1 for ψ reflects a situation with 90% certainty in the classification.

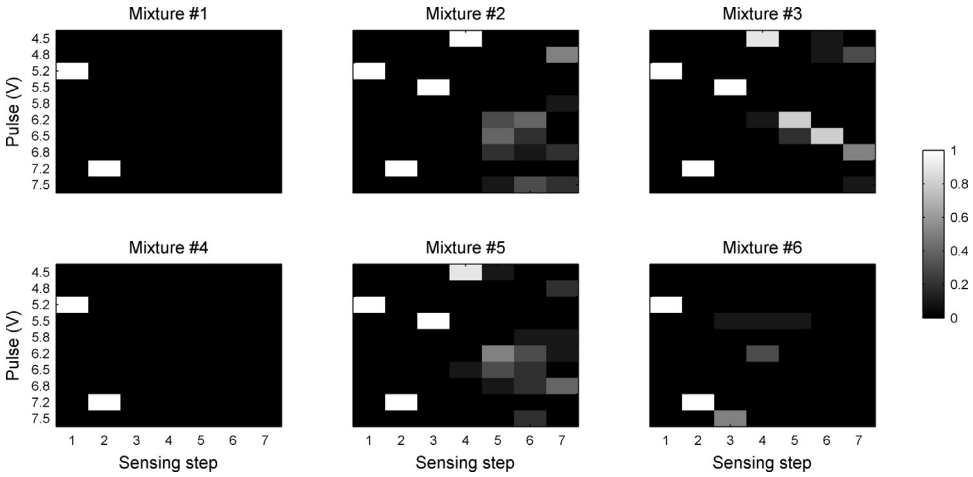

Fig. 10. Probability of selecting each temperature pulse as a function of the sensing steps for each mixture.

the sample. For the remaining mixtures (#2, #4, and #5), AS uses around 6–7 pulses. The low standard deviations in Table 3 suggest that the average number of pulses used by AS is very consistent with the composition of the test samples.

To illustrate the adaptive nature of AS, we also analyzed the probability that each pulse is selected as a function of the sensing step. Results for the case $f = 7$ (first round of cross-validation) are summarized in Fig. 10. In the first sensing step, AS always selects the same pulse (5.17 V), regardless of the test sample; since we assume a uniform belief distribution at $t = 0$, the selection in the first step is based solely on the sensor models. The same trend continues in the second sensing step where AS selects the same pulse (7.17 V) for all test samples. At the third sensing step, however, AS selects the pulse (5.5 V) for mixtures #2, #3, and #5, but selects a different pulse (7.5 V) for mixture #6. We observe similar behavior in the fourth sensing step: AS selects the pulse (4.5 V) for mixtures #2, #3, and #5, but the pulse (6.2 V) for mixture #6. In the latter sensing steps ($t \geq 5$), the sequences selected for mixtures #3 and #6 tend to converge to a single temperature, but for mixtures #2 and #5 AS select from among several of the higher temperatures. Thus, the results from Table 3 and Fig. 10 clearly illustrate the adaptive nature of AS, which allows it to select pulse sequences that are highly dependent on the identity of each sample.

5.4. Experiment IV: performance vs. pulse duration

Finally, we also investigated whether shorter pulse durations would affect the classification performance of the active sensing algorithm. For this purpose, we evaluated AS using only the first 15 s and 25 s of the transients as well as the entire 30-s transient. Following procedures described in Section 5.3, for each transient duration we first applied PCA to extract features, then used GMMs to create the sensor models, and finally estimated classification performance of AS using 10-fold cross validation at different cardinalities f.

Table 3
The average number of pulses used by AS and the corresponding standard deviations (at $f = 7$) for all the six mixtures. The mixture numbers correspond to those used in Table 2.

Mixture	1	2	3	4	5	6
Average # of pulses used	2.12	6.36	6.34	2.04	6.67	2.88
Standard deviation	0.16	0.53	0.38	0.08	0.24	0.68

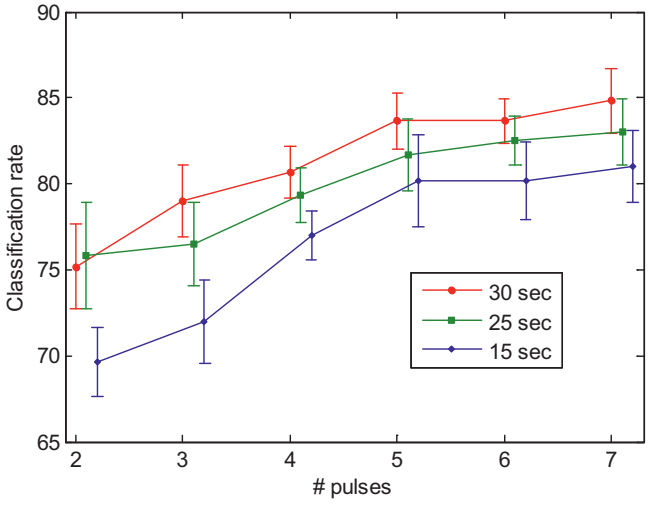

Fig. 11. Classification performance of AS vs. number of pulses used for three pulse durations – 15, 25, and 30 s. For visualization purposes, the curves corresponding to 15 s and 25 s pulses have been moved slightly along the x-axis.

Results are summarized in Fig. 11. As expected, classification rates are consistently higher for 30 s pulses than for 15 s or 25 s pulses, with the only exception of $f = 2$, where 30 s and 25 s pulses are comparable. The 30 s pulses improve the average classification performance by 1.3% and 4.5% points relative to 25 s and 15 s pulses, respectively, which indicates that the first half of the sensor transients contains most of the discriminatory information. This result is consistent with Fig. 8(a), which shows that the sensor often converges toward its steady-states after 15 s, rendering the latter parts of the transients relatively less informative. According to Fig. 11, differences between the short (15 s) and long (30 s) pulses are more marked with fewer sensing actions (2 or 3 temperatures), which suggests that accuracy can be improved by increasing either the number of sensing actions or the duration of the temperature pulses.

6. Conclusions and future work

We have presented an active-sensing approach for quantitative analysis of gas mixtures with MOX sensors. The approach allows a MOX sensor to adapt its operating temperature in real time to sequentially reduce uncertainty in the concentration estimates. We validated the approach on simulated and experimental data.

Our results indicate that active sensing is advantageous when either the sensing budget is limited or the sensor measurements are prone to noise.

The framework presented here could be extended to other types of temperature modulation waveforms. As an example, if the sensor temperature were to be modulated with sinusoidal waveforms [48] instead of pulses, AS could be used to choose the next modulation sine-wave (DC offset, amplitude, and frequency) based on information obtained from previous sensing steps. Examples of other temperature waveforms from recent literature where this framework could be used, include staircases [49], triangular waveforms [50], and voltage spikes [51]. In this paper, we used PCA for extracting features from the sensor transients. However, other feature extraction techniques such as wavelets [52], dynamic moments [48], or multi-exponential models [53] could easily be incorporated into the framework.

Identification times in our experiments are in the order of minutes, which may be impractical for certain applications. However, this is largely a limitation of the particular sensors we used for validation purposes (Figaro TGS sensors have long time-constants) and not of the active sensing method itself. Dramatically shorter response times can be obtained with newer metal-oxide sensor technologies, such as MEMS micro-hotplate arrays that have very low thermal constants, allowing for millisecond-scale temperature variations [21,54,55].

Our formulation is currently limited to discrete concentration spaces, i.e., where the concentration of each mixture component belongs to one of k possible discrete values. We are currently exploring ways in which the formulation could be extended to continuous concentration spaces, i.e., $\theta \in \mathbb{R}^n$. As an example, particle methods [56] can be used to search through a continuous concentration space. In this case, we would maintain a sample of L posterior distributions or particles $\theta_t = \{\theta_t^{[1]}, \theta_t^{[2]}, \ldots, \theta_t^{[L]}\}$, where each particle is an n dimensional vector representing a hypothesis of what the concentrations may be at time t, and each particle is assigned a weight $\omega_{t+1}^{[l]}$ which is proportional to the observational probability: $\omega_{t+1}^{[l]} \propto p(o_t|a_t, \theta_{t-1}^{[1]})$.

The myopic approach used in this paper (Table 1) selects temperatures based on their potential to reduce uncertainty in the belief distribution immediately, that is, following a sensing action. This is in contrast with exact solutions that are optimized over longer horizons though at additional computational costs. Thus, there is a trade-off between long-term and short-term gains [57], and the required computational complexity. However, considering the limited specificity of MOX sensors (even after temperature modulation) and the dynamic characteristics of chemical stimuli, it is unclear whether an optimal policy would provide significant improvements in performance when compared to the myopic approach tested in this paper.

In its present implementation, our myopic algorithm does not allow the same temperature to be measured multiple times. Because the sensor response is partly a function of its history, it is possible that measurements at the same temperature performed at different times could bring complementary information. Note, however, that in our experiments the sensor temperature is reset between consecutive temperature pulses (see Fig. 7). This removes, to a large extent, history effects on the sensor response, to where measurements of the same temperature at different times are likely to yield the same information. Multiple measurements at the same temperature may still be beneficial, e.g., as a way to increase signal-to-noise ratio [58]. However, the baseline method we used for comparison (SFS) does not allow repeated measures of the same feature, so it would have been at a disadvantage if we had allowed the active-sensing method to acquire multiple measurements at the same temperature. Leaving aside experimental considerations, our active sensing method can be easily extended to allow the same pulse multiple times.

Finally, our current implementation discretizes the observation space to estimate the expected entropy. This discretization step may be circumvented by using closed-form approximations of entropy. As an example, Huber et al. [59] have presented closed-form approximations of the true entropies for Gaussian mixtures, which are the exact form of our sensor models. Alternatively, one may estimate expected entropy via Monte Carlo sampling [60]. Monte Carlo sampling can provide more accurate estimates of differential entropy than discretization, but at a computational expense.

Acknowledgment

This work was supported in part by the National Science Foundation under Award #1002028.

References

[1] H.T. Nagle, S.S. Schiffman, R. Gutierrez-Osuna, The how and why of electronic noses, IEEE Spectrum 35 (1998) 22–31.
[2] A.P. Lee, B.J. Reedy, Temperature modulation in semiconductor gas sensing, Sensors & Actuators B: Chemistry 60 (1999) 35–42.
[3] C. Wang, L. Yin, L. Zhang, D. Xiang, R. Gao, Metal oxide gas sensors: sensitivity and influencing factors, Sensors 10 (2010) 2088–2106.
[4] S. Capone, A. Forleo, L. Francioso, R. Rella, P. Siciliano, J. Spadavecchia, D. Presicce, A. Taurino, Solid state gas sensors: state of the art and future activities, Journal of Optoelectronics and Advanced Materials 5 (2003) 1335–1348.
[5] S. Nakata, S. Akakabe, M. Nakasuji, K. Yoshikawa, Gas sensing based on a nonlinear response: discrimination between hydrocarbons and quantification of individual components in a gas mixture, Analytical Chemistry 68 (1996) 2067–2072.
[6] R. Gosangi, R. Gutierrez-Osuna, Active temperature programming for metal-oxide chemoresistors, IEEE Sensors Journal 10 (2010) 1075–1082.
[7] R. Gosangi, R. Gutierrez-Osuna, Quantification of gas mixtures with active recursive estimation, in: Proceedings of the ISOEN, 2011, pp. 23–24.
[8] K.D. Schierbaum, U. Weimar, W. Göpel, Multicomponent gas analysis: an analytical chemistry approach applied to modified SnO_2 sensors, Sensors & Actuators B: Chemistry 2 (1990) 71–78.
[9] H. Sundgren, F. Winquist, I. Lukkari, I. Lundstrom, Artificial neural networks and gas sensor arrays: quantification of individual components in a gas mixture, Measurement Science and Technology 2 (1991) 464–469.
[10] X. Wang, W.P. Carey, S.S. Yee, Monolithic thin-film metal-oxide gas-sensor arrays with application to monitoring of organic vapors, Sensors & Actuators B: Chemistry 28 (1995) 63–70.
[11] E. Llobet, R. Ionescu, S. Al-Khalifa, J. Brezmes, X. Vilanova, X. Correig, N. Barsan, J.W. Gardner, Multicomponent gas mixture analysis using a single tin oxide sensor and dynamic pattern recognition, IEEE Sensors Journal 1 (2001) 207–213.
[12] M.C. Burl, B.J. Doleman, A. Schaffer, N.S. Lewis, Assessing the ability to predict human percepts of odor quality from the detector responses of a conducting polymer composite-based electronic nose, Sensors & Actuators B: Chemistry 72 (2001) 149–159.
[13] A. Hierlemann, U. Weimar, G. Kraus, M. Schweizer-Berberich, W. Göpel, Polymer-based sensor arrays and multicomponent analysis for the detection of hazardous oragnic vapours in the environment, Sensors & Actuators B: Chemistry 26 (1995) 126–134.
[14] K. Domanský, D.L. Baldwin, J.W. Grate, T.B. Hall, J. Li, M. Josowicz, J. Janata, Development and calibration of field-effect transistor-based sensor array for measurement of hydrogen and ammonia gas mixtures in humid air, Analytical Chemistry 70 (1998) 473–481.
[15] W.P. Carey, S.S. Yee, Calibration of nonlinear solid-state sensor arrays using multivariate regression techniques, Sensors & Actuators B: Chemistry 9 (1992) 113–122.
[16] G. Huyberechts, P. Szecówka, J. Roggen, B.W. Licznerski, Simultaneous quantification of carbon monoxide and methane in humid air using a sensor array and an artificial neural network, Sensors & Actuators B: Chemistry 45 (1997) 123–130.
[17] C. Di Natale, F.A.M. Davide, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Göpel, A composed neural network for the recognition of gas mixtures, Sensors & Actuators B: Chemistry 25 (1995) 808–812.
[18] M. Pardo, G. Faglia, G. Sberveglieri, M. Corte, F. Masulli, M. Riani, A time delay neural network for estimation of gas concentrations in a mixture, Sensors & Actuators B: Chemistry 65 (2000) 267–269.
[19] S. De Vito, A. Castaldo, F. Loffredo, E. Massera, T. Polichetti, I. Nasti, P. Vacca, L. Quercia, G. Di Francia, Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures, Sensors & Actuators B: Chemistry 124 (2007) 309–316.

[20] A. Heilig, N. Bârsan, U. Weimar, M. Schweizer-Berberich, J.W. Gardner, W. Göpel, Gas identification by modulating temperatures of SnO_2-based thick film sensors, Sensors & Actuators B: Chemistry 43 (1997) 45–51.
[21] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, C. Cane, I. Gracia, X. Vilanova, X. Correig, Quantitative gas mixture analysis using temperature-modulated micro-hotplate gas sensors: selection and validation of the optimal modulating frequencies, Sensors & Actuators B: Chemistry 123 (2007) 1002–1016.
[22] R. Bajcsy, Active perception, Proceedings of the IEEE 76 (1988) 966–1005.
[23] L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, J.D. Schutter, Active sensing for robotics – a survey, in: Proceedings of the International Conference on Numerical Methods and Applications, 2002, pp. 316–324.
[24] R. Simmons, S. Koenig, Probabilistic robot navigation in partially observable environments, in: Proceedings of the IJCAI, 1995, pp. 1080–1087.
[25] H. Zhou, S. Sakane, Mobile robot localization using active sensing based on Bayesian network inference, Robotics and Autonomous Systems 55 (2007) 292–305.
[26] D. Fox, W. Burgard, S. Thrun, Active markov localization for mobile robots, Robotics and Autonomous Systems 25 (1998) 195–207.
[27] A.J. Davison, D.W. Murray, Simultaneous localization and map-building using active vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 865–880.
[28] K. Patel, W. Macklem, S. Thrun, M. Montemerlo, Active sensing for high-speed offroad driving, in: Proceedings of the IEEE ICRA, 2005, pp. 3162–3168.
[29] L. Pedersen, M. Wagner, D. Apostolopoulos, W.R. Whittaker, Autonomous robotic meteorite identification in Antarctica, in: Proceedings of the IEEE ICRA, 2001, pp. 4158–4165.
[30] A.O. Hero, D.A. Castañon, D. Cochran, K. Kastella, Foundations and Applications of Sensor Management, first ed., Springer, New York, 2007.
[31] Z. Yan, L. Xuejun, L. Carin, Detection of buried targets via active selection of labeled data: application to sensing subsurface UXO, IEEE Transactions on Geoscience and Remote Sensing 42 (2004) 2535–2543.
[32] S. Ji, R. Parr, L. Carin, Nonmyopic multiaspect sensing with partially observable markov decision processes, IEEE Transactions on Signal Processing 55 (2007) 2720–2730.
[33] J. Liu, P. Cheung, F. Zhao, L. Guibas, A dual-space approach to tracking and sensor management in wireless sensor networks, in: Proceedings of the ACM International Workshop on Wireless Sensor Networks and Applications, 2002, pp. 131–139.
[34] T.H. Chung, V. Gupta, J.W. Burdick, R.M. Murray, On a decentralized active sensing strategy using mobile sensor platforms in a network, in: Proceedings of the IEEE CDC, 2004, pp. 1914–1919.
[35] T. Nakamoto, N. Okazaki, H. Matsushita, Improvement of optimization algorithm in active gas/odor sensing system, Sensors & Actuators A: Physics 50 (1995) 191–196.
[36] C.E. Priebe, D.J. Marchette, D.M. Healy, Integrated sensing and processing decision trees, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 699–708.
[37] D.V. Dinakarababu, D.R. Golish, M.E. Gehm, Adaptive feature specific spectroscopy for rapid chemical identification, Optics Express 19 (2011) 4595–4610.
[38] R. Gutierrez-Osuna, A. Hierlemann, Adaptive microsensor systems, Annual Review of Analytical Chemistry 3 (2010) 255–276.
[39] C.H. Papadimitriou, J.N. Tsitsiklis, The complexity of markov decision processes, Mathematics of Operations Research 12 (1987) 441–450.
[40] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics, The MIT Press, Cambridge, 2005.
[41] D.M. Titterington, A.F.M. Smith, U.E. Makov, Statistical Analysis of Finite Mixture Distributions, Wiley, New York, 1985.
[42] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B 39 (1977) 1–38.
[43] J. Kittler, Feature set search algorithms, Pattern Recognition and Signal Processing (1978) 41–60.
[44] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1997) 273–324.
[45] P.K. Clifford, D.T. Tuma, Characteristics of semiconductor gas sensors. II. Transient response to temperature change, Sensors & Actuators 3 (1982) 255–281.
[46] S. Wlodek, K. Colbow, F. Consadori, Signal-shape analysis of a thermally cycled tin-oxide gas sensor, Sensors & Actuators B: Chemistry 3 (1991) 63–68.
[47] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
[48] A. Vergara, E. Llobet, E. Martinelli, C. Di Natale, A. D'Amico, X. Correig, Feature extraction of metal oxide gas sensors using dynamic moments, Sensors & Actuators B: Chemistry 122 (2007) 219–226.
[49] F. Hossein-Babaei, S.M. Hosseini-Golgoo, A. Amini, Extracting discriminative information from the Padé-Z-transformed responses of a temperature-modulated chemoresistive sensor for gas recognition, Sensors & Actuators B: Chemistry 142 (2009) 19–27.
[50] I. Montoliu, R. Tauler, M. Padilla, A. Pardo, S. Marco, Multivariate curve resolution applied to temperature-modulated metal oxide gas sensors, Sensors & Actuators B: Chemistry 145 (2010) 464–473.
[51] F. Hossein-Babaei, A. Amini, A breakthrough in gas diagnosis with a temperature-modulated generic metal oxide gas sensor, Sensors & Actuators B: Chemistry 166–167 (2012) 419–425.
[52] R. Ionescu, E. Llobet, Wavelet transform-based fast feature extraction from temperature modulated semiconductor gas sensors, Sensors & Actuators B: Chemistry 81 (2002) 289–295.
[53] R. Gutierrez-Osuna, A. Gutierrez-Galvez, N. Powar, Transient response analysis for temperature-modulated chemoresistors, Sensors & Actuators B: Chemistry 93 (2003) 57–66.
[54] B. Raman, J.L. Hertz, K.D. Benkstein, S. Semancik, Bioinspired methodology for artificial olfaction, Analytical Chemistry 80 (2008) 8364–8371.
[55] S. Semancik, R.E. Cavicchi, M.C. Wheeler, J.E. Tiffany, G.E. Poirier, R.M. Walton, J.S. Suehle, B. Panchapakesan, D.L. DeVoe, Microhotplate platforms for chemical sensor research, Sensors & Actuators B: Chemistry 77 (2001) 579–591.
[56] A. Doucet, V. Tadić, Parameter estimation in general state-space models using particle methods, Annals of the Institute of Statistical Mathematics 55 (2003) 409–422.
[57] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, 1998.
[58] T. Pearce, P. Verschure, J. White, J. Kauer, Robust stimulus encoding in olfactory processing: hyperacuity and efficient signal transmission, in: S. Wermter, J. Austin, D. Willshaw (Eds.), Emergent Neural Computational Architectures Based on Neuroscience, Springer-Verlag, Berlin, 2001, pp. 461–479.
[59] M.F. Huber, T. Bailey, H. Durrant-Whyte, U.D. Hanebeck, On entropy approximation for Gaussian mixture random vectors, in: Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 181–188.
[60] J. Denzler, C.M. Brown, Information theoretic sensor data selection for active object recognition and state estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 145–157.

Biographies

Rakesh Gosangi received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology Madras (IITM), Chennai, India, in 2007. He is currently working toward the Ph.D. degree in computer engineering at Texas A&M University, College Station. His research interests include pattern recognition, machine learning, active sensing, and machine olfaction.

Ricardo Gutierrez-Osuna received the B.S. degree in electrical engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1992, and the M.S. and Ph.D. degrees in computer engineering from North Carolina State University, Raleigh, in 1995 and 1998, respectively. He is currently a professor of computer engineering at Texas A&M University, College Station, TX. His current research interests include wearable physiological sensors, speech transformations for foreign accents, and active chemical sensing.