Odor recognition in robotics applications by discriminative time series modeling

**F.-M. Schleif, B.Hammer,
J. G. Monroy, J. Gonzalez-Jimenez, J. L. Blanco,
M. Biehl, N. Petkov**

Abstract Odor classification by a robot equipped with an electronic nose (e-nose) is a challenging task for pattern recognition since volatiles have to be classified quickly and reliably even in the case of short measurement sequences, gathered under operation in the field. Signals obtained in these circumstances are characterized by a high dimensionality, which limits the use of classical classification techniques based on unsupervised and semi-supervised settings, and where predictive variables can be only identified using wrapper or post-processing techniques. In this paper, we consider generative topographic mapping through time (GTM-TT) as an unsupervised model for time series inspection, based on hidden Markov models regularized by topographic constraints. We further extend the model such that supervised classification and relevance learning can be integrated, resulting in supervised GTM-TT. Then, we evaluate the suitability of this new technique for the odor classification problem in robotics applications. The performance is compared with classical techniques as nearest neighbor (NN), as an absolute baseline, support vector machine (SVM) and a recent time series kernel approach, demonstrating the eligibility of our approach for high dimensional data. Additionally, we exploit the learning system introduced in this work, providing a measure of the relevance of each sensor and individual time points in the classification process, from which important information can be extracted.

Keywords electronic nose, volatile classification, odor recognition, time series, prototype learning, relevance learning

1 Introduction

Olfaction plays an important role in the development of many applications, such as quality control in food processing chains, detection and diagnosis in medicine, finding drugs and explosives, and the more common estimation of blood alcohol content

Address(es) of author(s) should be given

(BAC) for drivers. Among them, there are some applications like pollution monitoring or leak detection that require to measure the environment continuously and at different locations. For such scenarios, the use of a mobile robot with the capability of identifying and measuring the volatiles' concentration is of great help as already reported in [26, 25]. Furthermore, olfaction also plays a key role in the development of more intelligent and useful robots at home, for example, by recognizing activities and environmental conditions, or improving social interaction [16].

Three are the main fields within robotics olfaction: gas distribution mapping (GDM) [4, 22], where the objective is to obtain a truthful representation of how volatiles are dispersed in the inspected area and their respective concentrations, gas source localization (GSL) where the robot is commanded to localize the emission sources [13], and odor recognition which deals with the problem of identifying which of a set of categories a new volatile sample belongs to [45].

The discrimination of gases performed with a robot equipped with an array of gas sensors presents a number of additional challenges when compared to standard analyte identification applications, mostly due to the differences in the measurement conditions. While standard classification tasks usually host gas sensors inside a chamber with controlled humidity, temperature and airflow conditions, in robotics olfaction there is no control over the sensing conditions. This entails that the sensor signals to be processed are noisy and dominated by the signal transient behavior.

Only few modeling methods are available to obtain interpretable, compact and precise predictive models for such type of data like [23, 7]. This is mainly due to the following reasons: (1) the number of time points is often low, while the dimensionality of the data is rather high, (2) the number of time sequences is often low, leading to a sparsely populated data space, (3) the sequences may have missing values, and may be of different length.

In this paper we demonstrate the suitability of a novel approach based on generative topographic mapping through time (GTM-TT) to the problem of volatile identification in robotics. The model extends classical GTM-TT by integrating supervised classification and relevance learning, resulting in supervised GTM-TT (SGTM-TT). More precisely, we have tested the SGTM-TT method with an e-nose comprising an array of MOX (metal oxide sensors) to classify samples of seven different volatiles under uncontrolled conditions. The performance is compared with techniques as nearest neighbor (NN), support vector machine (SVM) and a reservoir computing time series kernel (RTK) We illustrate one of the main advantages of the proposed method when classifying odors based on short data sequences, providing the predictive classification accuracy for sequences of reduced lengths (1s, 10s and 20s). Furthermore, we highlight the introduced relevance learning system for temporal high dimensional data, by studying the relevance of sensors and time points on the classification performance.

2 Related works

Odor discrimination with electronic noses has received growing attention and many studies have been done on how to classify odors using an array of gas sensors and

a pattern recognition algorithm. In [8,11,40,18] the principal methods for chemical classification with an array of gas sensors are reviewed, including nearest neighbor (NN), mahalanobis linear discriminant analysis, neural networks (ANN), cluster analysis with self-organizing Maps (SOM) and Support Vector Machines (SVM).

More recently, approaches based on ensembles of classifiers have been reported to improve the classification accuracy [44], improve the *earliness* on the classification [20], or to deal with the common problems of sensor drift and sensor replacement. In [47], a SVM based ensemble of classifiers is used to solve the gas discrimination problem over a period of three years by training different classifiers at different points of time. Similarly, in [27], a flexible classification strategy based on cooperative classifiers is proposed to increase the robustness of chemo-sensory systems against failures in their constituent sensing elements, postponing the necessity of replacing a sensor in the array, as well as facilitating the insertion of newly sensing elements.

Nevertheless, little attention has been given to the problem of classification in uncontrolled conditions, as revealed by the few works found in literature that perform classification focusing only on the transient phase of the sensor signals. An evaluation for the suitability of different feature extraction techniques for such scenarios is provided in [45], where Trincavelli et al. propose a preprocessing stage to isolate the relevant parts of the sensor signals that can then be passed to the pattern recognition algorithm. More recently, in [12] a Support Vector Machine is applied to a set of features obtained from changes of the spectral sensor signal characteristics (frequency components, phase shift and energy sums), reporting a substantially increase of the classification performance.

Gas sensor data has been analyzed by many different machine learning techniques with typically substantial preprocessing steps, limiting an out of sample extension, as discussed in more detail later on. Recent work [5] regarding the classification of gas sensor data is based on density estimates or models the time-series using decision trees [10].

Time series processing constitutes an advanced field of research with many powerful statistical analysis tools existing (see for example [41]). However, their methods usually require a sufficient length of the time series as compared to their dimensionality or consider only one-dimensional time series. Further the focus is often on modeling a timeseries, by means of a longer sequence to explore trends and predict future measurement values. In this work we are interested on discriminative models between different groups of time series and we would like to predict the class of the timeseries.

A few machine learning techniques exist to investigate high dimensional time series: Topographic mappings such as the self-organizing map (SOM) (see [1] for a recent review) were extended by a recursive context which accounts for the temporal dynamics [43]. A probabilistic counterpart is provided by the Generative Topographic Mapping Through Time (GTM-TT) which combines hidden Markov models with a constraint mixture model induced by a low dimensional latent space. This approach is extended to better take the relevance of the feature components into account in [31], but relying on an unsupervised model. The identification of relevant dimensions is very important as outlined e.g. in [31,23] to obtain a better understanding of the data,

to reduce the processing complexity, and to improve the overall prediction accuracy. A supervised relevance weighting scheme which singles out relevant features in a wrapper approach based on hidden Markov models has been proposed in [23]. In [7] a similar approach introducing class-wise constraints in the hidden Markov model is presented. In [23], applications to life science data are presented resulting in 85% prediction accuracy on a multiple sclerosis (MS) data set, but the approach makes multiple, restrictive assumptions regarding the used Hidden Markov Model (HMM). The approach [7] is evaluated in the same scenario with improved performance for the sclerosis data set. Ongoing work in the field reflects the high demand for effective methods for short but high dimensional time series data [33]. This is not limited to the bio-medical domain [23,7] but covers a broader field of applications in industry and geo-science [31,43]. In this work, we employ a supervised variant of GTM-TT (SGTM-TT) as introduced in [36] and extended in [37].

3 Method

3.1 Generative Topographic Mapping

As outlined before the complexity of the considered data requests for a strong regularizing and interpretable model. Topographic maps appear to be a good choice and especially the Generative Topographic Mapping (GTM) combines multiple necessary features. GTM was first introduced in [2] and models a given set of data vectors $\mathbf{x} \in \mathbb{R}^D$ in form of a mapping based on a constrained mixture of Gaussians. The mixture is induced by a lattice of points $\mathbf{w}$ in a low dimensional, so called, latent space which can also be used for visualization. The low dimensional lattice points are mapped by a projection $\mathbf{w} \mapsto \mathbf{t} = y(\mathbf{w}, \mathbf{W})$ into the high-dimensional data space. The corresponding mapping function is parametrized by the parameters $\mathbf{W}$; which usually are chosen in form of a generalized linear regression

$$y : \mathbf{w} \mapsto \varPhi(\mathbf{w}) \cdot \mathbf{W} \tag{1}$$

with basis functions $\varPhi$ as equally spaced Gaussians. The high-dimensional points $y(\mathbf{w}, \mathbf{W})$ are called prototypes and are determined in the original data space. The prototypes define a quantization of the original data space, representing the data with minimum possible error and can be inspected directly. For more recent work on prototype based learning and topographic maps see [1].

Every grid point of the GTM induces a Gaussian

$$p(\mathbf{x}|\mathbf{w}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left(-\frac{\beta}{2}\|\mathbf{x} - y(\mathbf{w}, \mathbf{W})\|^2\right) \tag{2}$$

with variance β^{-1}. Assuming a Dirac distribution of the prototypes, the data are modeled by a mixture of K modes

$$p(\mathbf{x}|\mathbf{W}, \beta) = \sum_{k=1}^{K} p(\mathbf{w}^k) p(\mathbf{x}|\mathbf{w}^k, \mathbf{W}, \beta) \tag{3}$$

multivariate (high-D) time series   *low-D 'states'*

high-D trajectory   transition probabilities
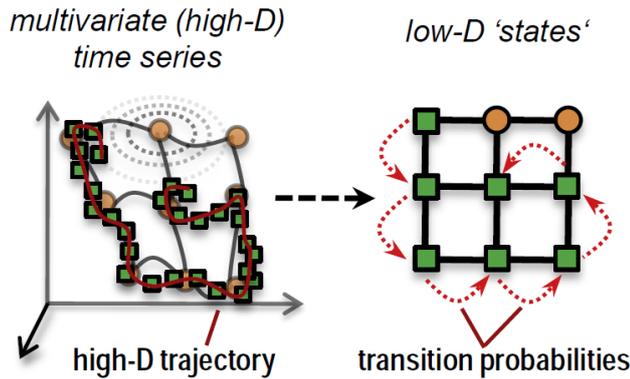
Fig. 1: GTM-TT consisting of a hidden Markov model, which hidden states are constrained to be organized on a grid topology (the latent points of the GTM model). The emission probabilities are governed by the GTM mixture distribution [2]. In the left figure a data distribution is given in a 3D space with an intrinsic low-dimensional support. Additionally these data are not i.i.d. but dependent over time leading to some trajectory. GTM is used to project the data to a low dimensional grid (here 2D, right plot). The prototypes (circles left) are generated by the latent points (in 2D, right) as HMM constrained Gaussians (left, dotted circles). Here we consider 9 hidden states organized on a 3×3 grid. The data distribution may change over time and hence also the mapping of the GTM is effected over time, assuming smooth transitions within the HMM.

with $p(\mathbf{w}^k) = 1/K$, assuming equal probabilities of the modes. We optimizes the data log-likelihood

$$\ln \left(\prod_{n=1}^{N} \left(\sum_{k=1}^{K} p(\mathbf{w}^k) p(\mathbf{x}^n | \mathbf{w}^k, \mathbf{W}, \beta) \right) \right) \tag{4}$$

by means of an expectation maximization (EM) strategy with respect to the model parameters $\mathbf{W}$ and β with data dimensionality D and number of points N as detailed in [2].

Finally an unsupervised restricted Gaussian mixture model (GMM), induced by a low dimensional latent space, is defined.

3.2 GTM Through-Time

For temporal data the original GTM formulation is limited because it does not account for the dependency between different time points leading to quite complex and redundant GTM models (see [31]). An extension was provided by the GTM through time (GTM-TT) [2] where the entries over time are no longer independent. It basically provides an advanced time-series clustering using a constrained hidden Markov model, which is useful under our given constraints. It is assumed that the data are
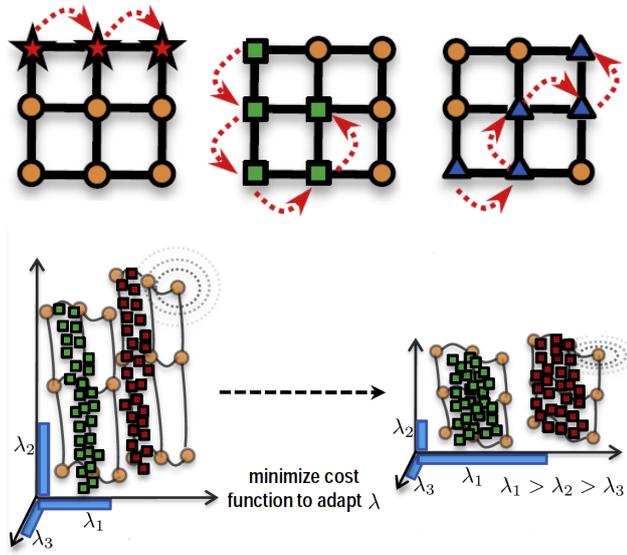
Fig. 2: Illustration of the SGTM-TT. It consists of multiple GTM-TT models. It behaves similar to the regular GTM-TT but the training is classwise and the β parameter is common over the different models. The different classwise models are used to represent the data distribution over time (here for three classes). In the bottom the SGTM-TT with relevance learning is shown. The relevance of the input-dimensions is weighted over time during training. And only relevant dimensions with large λ-values are kept. In the figure the Λ_1 dimension discriminates the two groups and is pronounced by metric adaptation.

time series in the D-dimensional metric space, i.e. $\mathbf{x} = \mathbf{x}(1)\ldots\mathbf{x}(T) \in (\mathbb{R}^D)^*$ where $T \geq 1$ is the length of the time series. A data point of the training data will be referred to as $\mathbf{x}^i$. We assume that entries, consecutive in time, $\mathbf{x}(t)$ and $\mathbf{x}(t+1)$ are strongly correlated. In the GTM-TT the observation space (over time) is represented by a topographic mapping as described before but its time dependence is modeled in form of a hidden Markov model (HMM). In the GTM-TT model the hidden states are given by the lattice points $\mathbf{w}^j$. The concept of the GTM-TT is depicted in Figure 1. Lets assume a given sequence $\mathbf{x}$ of observations and an underlying sequence of hidden states of the same length $\mathbf{z} = \mathbf{z}(1)\ldots\mathbf{z}(T)$ where $\mathbf{z}(i)$ is equivalent to a point $\mathbf{w}^j$. Then, the probability of the observations and a corresponding path of hidden states $\mathbf{z}$ can be described by $p(\mathbf{x}, \mathbf{z}|\Theta) =$

$$p(\mathbf{z}(1))\prod_{t=2}^{T} p(\mathbf{z}(t)|\mathbf{z}(t-1), \mathbf{W}, \beta)\prod_{t=1}^{T} p(\mathbf{x}(t)|\mathbf{z}(t)) \tag{5}$$

with the conditional probability $p(\mathbf{x}(t)|\mathbf{z}(t)) := p(\mathbf{x}(t)|\mathbf{z}(t), \mathbf{W}, \beta)$ is as before (2) [2]. This results in the overall probability of $\mathbf{x}$: $p(\mathbf{x}|\Theta) = \sum_{\mathbf{z}\in\{\mathbf{w}^1,\ldots,\mathbf{w}^K\}^T} p(\mathbf{x}, \mathbf{z}|\Theta)$

For the parametrization of the GTM-TT ($\Theta = (\mathbf{W}, \beta, \pi, \mathbf{P})$) we rely on the assumption of the standard Markov property and stationarity of the dynamics. With additional parameters for the initial state probabilities $\pi = (\pi_j)_{j=1}^K$ where $\pi_j = p(\mathbf{z}(1) = \mathbf{w}^j)$ and transition probabilities $\mathbf{P} = (p_{ij})_{i,j=1}^K$ where $p_{ij} = p(\mathbf{z}(t) = \mathbf{w}^j | \mathbf{z}(t-1) = \mathbf{w}^i)$, the latter one characterizing the temporal correlations of subsequent states. The data log likelihood is optimized by: $\ln\left(\prod_{n=1}^N p(\mathbf{x}^n|\Theta)\right)$. using an EM-approach. Like for standard HMMs the hidden parameters (responsibilities) are defined by a forward-backward procedure [48]. Based on these parameters $\mathbf{W}$ and β can be determined as specified before. The probability of being in state $\mathbf{w}^k$ at time t, given the observation sequence $\mathbf{x}^n$ (responsibilities) is given as:

$$r^{kn}(t) = p(\mathbf{z}(t) = \mathbf{w}^k | \mathbf{x}^n, \Theta) = \frac{A_{kt} B_{kt}}{p(\mathbf{x}^n|\Theta)} \qquad (6)$$

Using the joint probability $p(\mathbf{x}^n(1) \ldots \mathbf{x}^n(t), \mathbf{z}(t) = \mathbf{w}^k | \Theta)$ and the subsequent equation:

$$A_{kt} = \sum_{i=1}^K A_{it-1} p_{ik} p(\mathbf{x}^n(t)|\mathbf{w}^k, \Theta) \qquad (7)$$

we get the forward variable A_{kt} with the start condition $A_{k1} = \pi_k p(\mathbf{x}^n(1)|\mathbf{w}^k, \Theta)$. The variable B_{kt} is the joint probability $p(\mathbf{x}^n(t+1) \ldots \mathbf{x}^n(t_n), \mathbf{z}(t) = \mathbf{w}^k | \Theta)$ and is calculated using $B_{kt} = \sum_{i=1}^K p_{ik} p(\mathbf{x}^n(t+1)|\mathbf{w}^i, \Theta) B_{it+1}$ where $B_{kT} = 1, B_{kt}$ defines the backward variable. The transition parameters are trained using the standard Baum-Welch training. As usual the underlying HMM also permits to deal with missing values and sequences of arbitrary length [3]). A more detailed description of the GTM-TT is given in [42].

For an input time series $\mathbf{x}^n(1) \ldots \mathbf{x}^n(T)$, GTM-TT specifies a time series of responsibilities $r^{kn}(1) \ldots r^{kn}(T)$ of neuron k. This can be used to define a winner for every time step t: $\operatorname{argmax}_k r^{kn}(t)$.

3.3 Supervised GTM-TT

In the considered problem scenario our time series data provide additional label information, such that $\mathbf{x}$ is equipped with a label l, element of a finite label set $\{1, \ldots L\}$. We also assume that the given label is constant over time. Now, we would like to incorporate the label information in the optimization process of the GTM-TT leading to an extended supervised classification scheme. Given a labeled training set, we learn a separate GTM-TT for every class, whereby the models are linked by the same bandwidth β and the same underlying topological grid. We also use the same basis functions Φ and the Dirac distribution on the latent space. However, the prototype parameters $\mathbf{W}_l$, the initial state probability π_l and the transition probabilities $\mathbf{P}_l$ are learned individually for every model representing label l. We refer to this model as the Supervised GTM-TT (SGTM-TT) as depicted schematically in Figure 2. Accordingly, we will have a quantitative model for every class l after training.

In the recall or test phase we have to analyze a novel time series $\mathbf{x}$ and obtain L time series of predicted responsibilities according to every model which will be denoted by $r_l^k(\mathbf{x}(t))$ (responsibilities of model l for input $\mathbf{x}$ at time point t). We can summarized the responsibilities in an aggregated form as:

$$r_l(\mathbf{x}) := \sum_{k=1}^{K} \sum_{t=1}^{T} r_l^k(\mathbf{x}(t))/(KT) \tag{8}$$

and one can select the label l as predicted output for which this value is largest.

3.4 Relevance learning for SGTM -TT

Metric adaptation for discriminative prototype based learning has been introduced in [19], it is often also referred to as relevance learning. The basic idea is to parametrize the distance measure to incorporate auxiliary information. For the squared Euclidean metric one can define a parametrized, weighted, variant:

$$d_{\boldsymbol{\lambda}}(\mathbf{x}, \mathbf{t}) = \sum_{d=1}^{D} \lambda_d^2 (x_d - t_d)^2 \,. \tag{9}$$

For the GTM such a parametrization was already discussed in [15] for i.i.d. data resulting in relevance GTM (R-GTM). However, having temporal data some adaptations are necessary and also the supervision has to be handled in an alternative way. To keep the approach simple and to limit the number of free parameters we will restrict our approach to a global diagonal weighted distance, in which case a weight λ_i directly corresponds to the relevance of dimension i. Here we assume normalized data with mean 0 and a standard deviation of 1 for each dimension. For GTM-(TT), the distance used to compute local probabilities is replaced by the previously discussed weighted Euclidean distance:

$$p_{\lambda}(\mathbf{x}|\mathbf{w}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left(-\frac{\beta}{2} d_{\lambda}(\mathbf{x}, y(\mathbf{w}, \mathbf{W}))\right) \tag{10}$$

Accordingly the data log likelihood considers the relevance of the data dimensions and, hence we obtain a corresponding topographic mapping.

A main difference of this approach to a standard integration of a data correlation matrix into the Gaussians consists in the fact that we prefer to adapt the relevance parameters in a supervised way according to the given label information, resulting in a discriminative approach.

The relevance parameters λ are optimized as suggested in [15] using the class information in an additional update step, interleaved with the standard adaptation of the SGTM-TT using the parametrized distance.

The discriminative learning of the metric parameters is controlled by the cost function of the generalized learning vector quantization (GRLVQ) which is a large margin technique [39]. We assume a classification based on a finite set of prototypes

$\mathbf{t}^j$ which are equipped with class labels and represent the given data. A classification is done by means of a winner takes all scheme: the predicted label corresponds to the prototype with smallest distance $d_\lambda(\mathbf{x}, \mathbf{t}^j)$. For standard GTM, our prototypes are given by latent points $\mathbf{t}^j = y(\mathbf{w}^j, \mathbf{W})$, and the distances determine the responsibilities of the data points. The relevance terms λ are adapted such that the costs

$$E(\lambda) = \sum_n \mathrm{sgd}\left(\frac{d_\lambda(\mathbf{x}^n, \mathbf{t}^+) - d_\lambda(\mathbf{x}^n, \mathbf{t}^-)}{d_\lambda(\mathbf{x}^n, \mathbf{t}^+) + d_\lambda(\mathbf{x}^n, \mathbf{t}^-)}\right) \tag{11}$$

are minimized. The closest prototype with the correct labeling is denoted by $\mathbf{t}^+$ and the one with the incorrect label by $\mathbf{t}^-$, for a given input $\mathbf{x}^n$. The sigmoid function (sgd) is defined as: $sgd(x) = \frac{1}{1+\exp(-\sigma \cdot x)} \in [0,1]$ This optimization scheme can be integrated into the vectorial GTM, simultaneously adapting the GTM parameters, optimizing the data log-likelihood, and the metric parameters optimizing the classification margin. The update equations for the parameters λ can be derived from (11), taking the derivatives. To keep a quadratic form in the distance measure, the metric parameters are normalized after each adaptation step.

Given an input sequence $\mathbf{x}$ we get a prototype representation of this time series by evaluating the SGTM-TT in the following way. For every class label we consider the time series of prototypes of the corresponding GTM-TT model according to the winner prototypes over time:

$$\mathbf{t}_l = (\mathbf{t}_l(1) \ldots \mathbf{t}_l(T)) \tag{12}$$

where

$$\mathbf{t}_l(t) = y(\mathbf{w}^k, \mathbf{W}_l) \text{ with } k = \mathrm{argmax}_k r_l^k(\mathbf{x}(t)) \tag{13}$$

Now the time series $\mathbf{x}$ and the corresponding time series of prototypes representing a correct or a wrong class label can be used in (11) to adapt the underlying metric. If we assume an appropriate metric for the comparison of two time series, a well defined cost function results.

Several reasonable distance measures for time series can be considered, whereby the only property which we will use is differentiability. For simplicity we will also assume, that the time series have equal length, although the model can be generalized to time series of different length.

A very simple distance for such time series would be to average over the Euclidean distances in each time point. This however is inappropriate, because it will completely neglect the functional form of the data. An appropriate measure, designed for the comparison of timeseries was proposed in [21] and will be used instead. Further alternatives time series metrics are possible see e.g. [10], but the chosen one has been found to be effective in prior work [38] and can be calculated at low costs.

The considered distance measure integrates the functional form of three subsequent time steps in comparing $\mathbf{x}(t)$ and $\mathbf{t}(t)$. Let us assume we have a real valued time series $\mathbf{v} = v(1) \ldots v(T)$, then the functional L_p norm can be defined as [21]:

$$\mathcal{L}_p^f(\mathbf{v}) = \left(\sum_{t=1}^T (\triangle A_t(\mathbf{v}) + \triangle B_t(\mathbf{v}))^p\right)^{\frac{1}{p}} \tag{14}$$

with

$$\triangle A_k(\mathbf{v}) = \begin{cases} \frac{\tau}{2}|(t)| & \text{if } 0 \leq v(t)v(t-1) \\ \frac{\tau}{2}\frac{(t)^2}{|v(t)|+|v(t-1)|} & \text{if } 0 > v(t)v(t-1) \end{cases} \tag{15}$$

$$\triangle B_k(\mathbf{v}) = \begin{cases} \frac{\tau}{2}|v(t)| & \text{if } 0 \leq v(t)v(t+1) \\ \frac{\tau}{2}\frac{v(t)^2}{|v(t)|+|v(t+1)|} & \text{if } 0 > v(t)v(t+1) \end{cases} \tag{16}$$

representing the triangles on the right and the left sides of $\mathbf{v}(t)$ and boundary points are set to 0. This norm accounts for entries which change the sign in subsequent time steps. We obtain a weighted distance, for vectorial data $\mathbf{x}$ and $\mathbf{t}$ over time with equal dimensionality D at each time point:

$$d_\lambda(\mathbf{x}, \mathbf{t}) = \sum_{i=1}^{D} \lambda_i \mathcal{L}_p^f(\mathbf{x}_i - \mathbf{t}_i) \tag{17}$$

where $\mathbf{x}_i - \mathbf{t}_i$ refers to the time series of real numbers given by the distance of the entries in dimension i. As a special property of this distance measure the similarity of the curvature of the sequences is taken into account. Again, each dimension is weighted by the normalized relevance parameters λ_i.

This weighted metric (17) is used in the cost function (11). If we take the derivatives (see [38] for $\mathcal{L}_p$-norm) with respect to the relevance terms an adaptive weighting for the input dimensions is obtained taking the functional form of the data into account. Again the λ are normalized after every adaptation to obtain non-negative values, summing up to 1.

Relevant time points:

Since SGTM-TT relies on HMMs, every time point depends on its predecessor only. Thus, it is not reasonable to adapt the relevance of time points to obtain a better representation of data in the GTM-TT models. However, it is reasonable to judge the relevance of time points resulting from the GTM-TT models for the final classification, in particular if time series are of the same or a similar length. This method offers insights into the model to identify time points which are particularly discriminative for the given task at hand.

We obtain a relevance profile in the following way: Denote by $r_l(\mathbf{x}(t)) := \sum_{k=1}^{K}(r_l^k(\mathbf{x}(t)))/K$ the accumulated responsibility of the GTM-TT model l for data point $\mathbf{x}^n$ at time point t. Based on this value, a classification can be based on the maximum responsibility $r_l(\mathbf{x}(t))$ in time point t. For every time point t, we simply count the number of data points which are classified correctly as belonging to class l based on the classification for time point t only, averaged over all data. A global relevance profile results thereof as a sum over all labels.
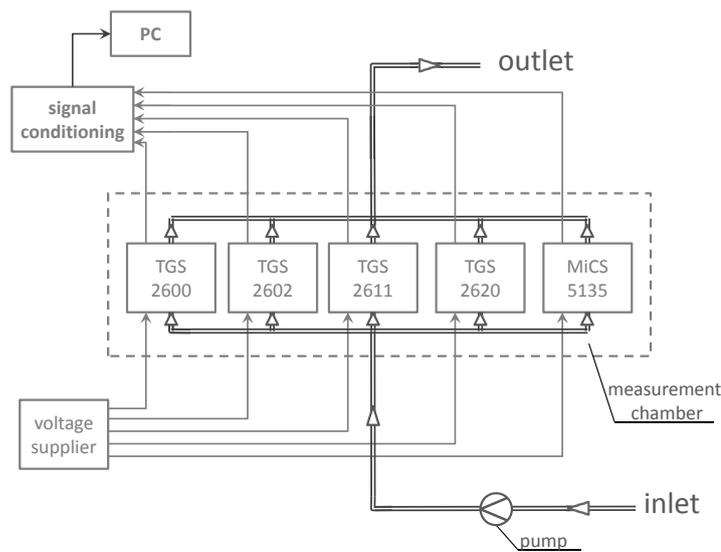
Fig. 3: Measurement system.

4 Odor Measurement System

The analyte measurement system employed to gather the data presented in this article is shown in Figure 3. It consists of an array of metal oxide semiconductor (MOX) gas sensors hosted inside a measurement chamber, a pneumatic circuit to control the exposition of the sensors to the volatile molecules dispersed in the environment, and the electronics necessary to power up the sensors and respective measurement circuits.

The election of MOX as the gas sensing technology has been made attending at its high sensitivity, commercial availability and low price. However, they present some shortcomings including: poor selectivity, influence by environmental factors such as humidity and temperature [30] and major limitations in their response speed [29]. Among these drawbacks, their poor selectivity is of the largest concern for odor classification. To overcome this, it is a common practice to build the e-nose upon an array of MOX sensors with different and partially overlapping sensitivities. The output of the array is then processed with a pattern recognition algorithm to find out which substance the e-nose is exposed to. Based in this concept, we choose five different MOX gas senors to compose the sensor array: TGS-2600, TGS-2602, TGS-2611 and TGS-2620 from Figaro Sensors[1], and MiCS-5135 from e2V Sensors[2].

In order to enable sensors to interact with the volatile molecules dispersed in the environment, the e-nose employs a pump to enforce a constant airflow through the sensors array. The aspiration and release of the air samples are accomplished through tubes, conveniently separated one to another to avoid cross contamination. Addition-

---

[1] Figaro engineering inc. http://www.figaro.co.jp

[2] e2v. http://www.e2v.com/.

ally, the aspiration through flexible tubes allows the displacement of the aspiration entry without the need to move the complete system. This advantage is particularly useful in robotics to easily sample the space, for example, by attaching the e-nose aspiration to the hand of an arm robot as shown in the experimental section.

4.1 Signal conditioning and data preprocessing

The data, as provided by the e-nose, present a measurement intrinsic baseline, which can be seen as a signal offset. Here, we estimate the baseline value as the median signal intensity within the first $5 - 20$ seconds, and then, remove it from each measurement truncating values to zero when necessary.

More sophisticated preprocessing, by means of advanced baseline correction algorithms, smoothing strategies or normalization techniques [35] are possible but out of focus of this paper. We also do not further explore specific feature extraction techniques for spectral data but focus on the obtained normalized intensities.

5 Experimental Results

This section describes the setups and classification results for three different experiments designed with increasing classification challenge. Furthermore, a comparison of results with SVM, NN and a very recent reservoir computing based time series classifier (RTK) as proposed in [6] is provided. For RTK the core idea is to transform the time series into a higher dimensional dynamical feature space via reservoir computation models. Subsequently varying aspects of the signal are represented through variation in the linear readout models trained in such dynamical feature spaces, for details see [6].

In general we are interested on simple methods or at least methods which provide direct interpretation of the model parameters and results. For example it is very desirable to have direct links to the input features to find channels which are most discriminative for a specific substance, relevant over all classes but also the other way, being not very relevant. The later is an important characteristic for systems with limited resources, like mobile robotics, where it would be desirable to power on only the relevant sensors. Accordingly (local) linear methods are interesting in contrast to black box non-linear kernel mappings. We are also interested on approaches which permit an easy and quick out of sample extension to, in our case, substantially shorter sequences in the test phase. This rules out multiple complicated time series models.

For SVM we used a linear kernel with optimal C determined over the training data on a grid search. Since SVM can not directly be applied to temporal data, nor can it be used for sequences of different length in a direct way, for the comparison we simply concatenate the measurements of the different channels to remove the time dimension. More complex strategies of applying SVM, e.g. by using a dynamic time warping (DTW) kernel could be done but are not in the focus of this paper and out of sample extensions are often not immediate which is an issue for online robotic sensor systems. For more recent work around DTW or kernel related time series analysis

see e.g. [32, 28, 6]. Additionally we would like to avoid more complex preprocessing steps to permit an easy out of sample extension in practical settings. Although maximum classification performance is not our main objective, we also provide a comparison with a very recent reservoir computing kernel [6]. This approach is known to be very effective for timeseries but on the other hand is less interpretable nor is the out of sample extension for very short sequences immediate. For RTK there are three parameters optimized on the training data within a grid search[3] as detailed in [6].

5.1 Experiment 1: Simulated data

The first experiment is based on the simulated data proposed in [23] with the only intention of validating the proposed algorithm under known conditions.

The simulated data (SIM) consist of 100 samples separated into two classes of 50 samples each. Each point is located in a 100 dimensional feature space with 8 time points. From the given features, only 10 are expected to differentiate between the classes. Details about the data and the generation procedure are given in [23].

We applied SGTM-TT with relevance learning using 9 hidden states and 4 basis functions. We observe an overall prediction accuracy of $94 \pm 4\%$. The relevance profile identified all known 10 features and effectively pruned out the remaining irrelevant data dimensions. Our results are slightly better than those reported in [23] (90%) and in [7] (92%).

The dataset is a particular short time series with a rather large number of input dimensions. Especially the small number of time points can be quite challenging for other time series models but may actually occur in the context of electronic nose experiments, where short sensing cycles would be very desirable. The prediction results of the different methods are summarized in Table 1. With the exception of NN most methods perform reliable well but SGTM-TT was significantly better.

5.2 Experiment 2: Controlled gas exposure

The second experiment aims to test the proposed method with real odor data under restrained environmental conditions. To this end, a dataset of real odor samples is gathered in a scenario as controlled as possible. The dataset is comprised by 39 samples generated by exposing the e-nose to gas pulses of four different analytes: a commercial spirit (*Larios Gin*), a polish remover based on Acetone, standard ethanol and lighter gas (butane mixed with propane). Acetone was given by 9 samples and the other classes by 10 samples each.

Each sample is collected according to the following three-phases procedure: (1) for the initial 30s, baseline value is estimated by measuring the sensor response in absence of the target gas, (2) then, for a duration of 60s the e-nose is placed next to the gas source (about 10cm) exposing the sensor array to the volatile. Finally (3),

---

[3] Grids: $\lambda, \gamma = [0, 10^{-6} \dots 10^{-1}, 0.5, 1 \dots 5, 10, 30, 50, 100]$ $costs = [0.1, 10, 10^2, 5 \cdot 10^2, 10^3, 5 \cdot 10^3, 10^4, 5 \cdot 10^4]$

409 the gas source is removed allowing the sensor array to recover to its initial state
410 (baseline).

411   Figure 4 shows two different samples of such dataset. Notice that although the
412 gas exposure was "controlled" by time exposure and distance to the source, strong
413 fluctuations in the sensor readings occur due to the chaotic nature of the gas disper-
414 sion.

415   The SGTM-TT is inherently capable of dealing with measurement sequences of
416 different length in time, using the HMM mapping functionality. However, to permit
417 fair comparison with other approaches like vector embeddings, we consider only the
418 first 100 sec. of the data. That is, we built a first dataset (DS1) using the initial 100sec.
419 of each sample, which corresponds to 487 sampling points.

420   For comparison we also use two public domain data sets of similar type (elec-
421 tronic nose data) from the UCI database. The DS-UCI-1 data set is given by the
422 *two sources gas* data [14]. The data are measured using a chemical detection plat-
423 form composed of 8 chemo-resistive gas sensors which were exposed to turbulent
424 gas mixtures generated naturally in a wind tunnel. It consists of 180 time series of
425 Ethylene (Eth) ,Carbon Monoxide (CO) and Methane (Me) mixtures at different con-
426 centrations. We use the data as a two class prediction problem to predict the whether
427 Eth was mixed with CO or Me. Available features are temperature, humidity and the
428 8 sensor channel outputs. Each time series is given with 2970 sampling points.

429   The DS-UCI-2 data set is given by the *pulmon* data [49]. The data are measured
430 using a chemical sensing system based on an array of 16 metal-oxide gas sensors and
431 an external mechanical ventilator to simulate the biological respiration cycle. The
432 tested gas classes are mixtures of acetone and ethanol. Data have been normalized
433 to zero-mean and intensity and considered again as a prediction problem to identify
434 whether the mixture contains Me or CO.

435 The classification accuracy for DS1, DS-UCI-1, DS-UCI-2 is given in Table 1 in
436 comparison to some standard approaches. We observe that the SGTM-TT performs
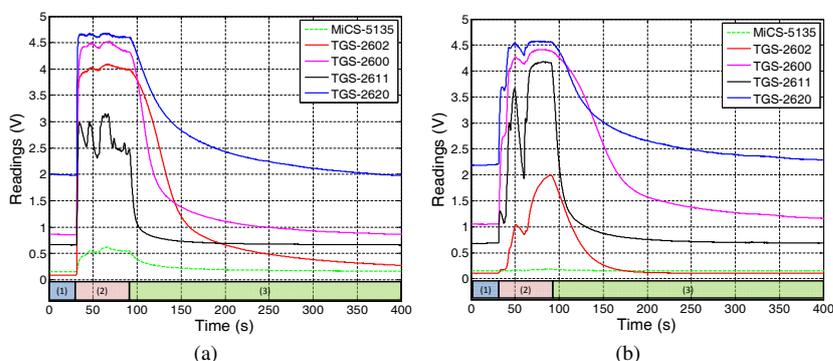
Fig. 4: Two different samples of the olfaction dataset gathered in the second experi-
ment. The three phases in which the samples can be decomposed are marked at the
bottom of each figure as (1),(2) and (3).

CV-Accuracy	SGTM-TT	SVM	NN	RTK
SIM	$\underline{\mathbf{94.00}} \pm 4.18\%$	$90.00 \pm 5.00\%$	$55.00 \pm 13.54\%$	$66.30 \pm 8.54\%$
DS1	$88.03 \pm 9.72\%$	$86.36 \pm 9.66\%$	$80.49 \pm 11.90\%$	$\underline{\mathbf{96.67}} \pm 4.56\%$
DS-UCI-1	$87.78 \pm 5.76\%$	$\underline{\mathbf{93.89}} \pm \mathbf{4.97}\%$	$86.81 \pm 7.98\%$	$64.44 \pm 4.12\%$
DS-UCI-2	$79.55 \pm 9.15\%$	$83.03 \pm 18.47\%$	$76.33 \pm 18.15\%$	$\underline{\mathbf{94.70}} \pm 8.05\%$

Table 1: Average test set accuracy for the first and second experiment in a 5 fold cross-validation. Significant better results are underlined.
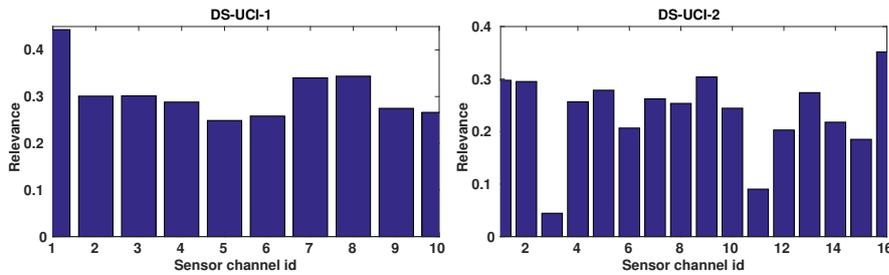
Fig. 5: Relevance profile of the sensor input for DS-UCI-1 (left) and DS-UCI-2 (right). For both profiles the information is distributed over the various sensors but some sensors are more important e.g. sensor 1 for DS-UCI-1 and sensor 3 for DS-UCI-2.

reliable well although the best prediction accuracy for DS1 and DS-UCI-2 is obtained by the RTK approach. For the DS-UCI-1 dataset RTK is significantly worse than the other approaches and the SVM obtained the best performance. Hence there is not a clear winner regarding the classification accuracy but SGTM-TT represents a good approach with a reliable and consistent performance. Furthermore, as previously commented, the classification performance is not the only point that matters but also the simplicity of the model and the interpretability of the results. Neither RTK nor SVM provide additional insight in the relevance of the sensor channels [4]. Here we are mainly interested on interpretable models [24] which also simplify a later transfer of the approach to an embedded system or the sensor platform. In Figure 5 we show the averaged (global) sensor relevance profile of DS-UCI-1 and DS-UCI-2.

Subsequently we give a detailed analysis for our own dataset - DS1, where we have more background information to provide a specific in depth discussion of the results. For the analysis of the sensor relevance and time points relevance, the whole measurement sequence of each sample was down-sampled to 800 time points each (DS2). The SGTM-TT was then trained in a 5-fold crossvalidation with 4 hidden states and 4 basis functions. In Figure 6 we show the relevance indexes of the five gas sensors of the e-nose for the different target volatiles of DS2 as obtained by SGTM-TT. Different conclusions can be drawn from the study of such relevance plot:

---

[4] Approaches for feature ranking by SVM are available but not for this type of data and not directly for multi-class problems as studied for DS1.

456 – In general, the five MOX sensors are relevant for the classification of the differ-
457   ent volatiles, being sensor $TGS - 2620$ the less relevant one, and so the most
458   expendable.
459 – Sensor $TGS - 2602$ is the most relevant one when classifying Acetone and
460   Ethanol samples, with a notable difference with respect the other sensors in the
461   case of Acetone. This characteristic is already reported in the manufacturer's
462   datasheet, indicating the high sensitivity to volatile organic compounds (VOCS)
463   of this sensor model.
464 – As expected from the low selectivity characteristic of MOX sensors, each sensor
465   presents a high relevance index for more than one odor class.

466 We also explore the relevance of individual time points of the dataset DS2, depicted
467 in Figure 7(a). As expected, the time-interval under volatile exposition, the first 100
468 seconds, is the most discriminating. Furthermore, and as already reported in [9], it is
469 noticeable the fact that relevant information for classification purposes can be found
470 in the recovery phase, after the volatile has been removed.
471   Since in real robotics conditions the classifier is expected to work on small data
472 sequences, a second configuration for the dataset DS2 was tested. Here, the test data
473 consist only of short sensor readings over time. Figure 7(b) depicts the accuracy in the
474 classification for three different window lengths (1s, 10s and 20s). We observe that
475 given the highly dynamic response of MOX sensors in addition to the inherent sig-
476 nal noise, very small windows (1s) do not carry enough information for a reasonable
477 classification, but for data sequences of ten seconds the accuracy in the prediction
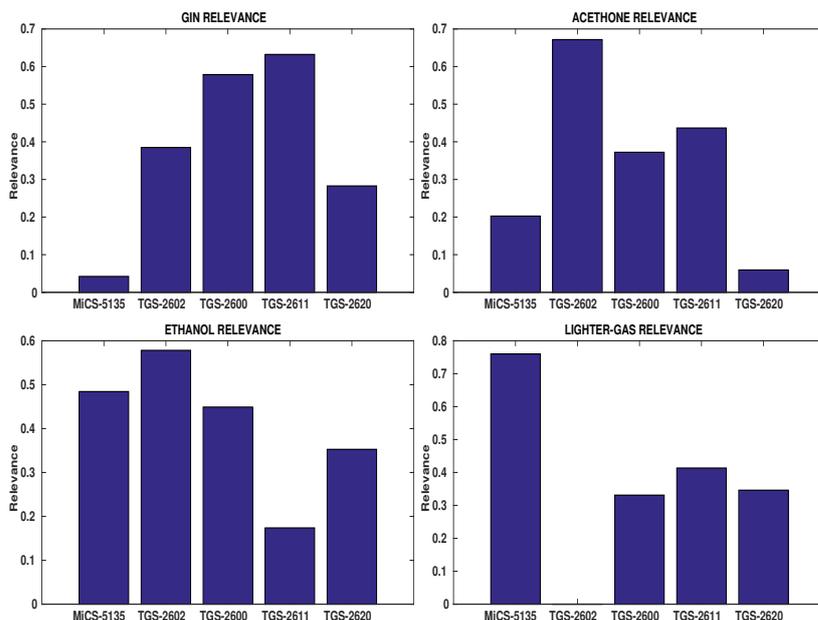
Fig. 6: Sensor relevance indexes for the four odor classes used on dataset DS2.

achieves very good results (values near 0.8). Furthermore, window lengths over ten seconds seems to not improve the accuracy, which indicates that long sequences encode a lot of noise contributions, hampering the model in the prediction. Finally, it must be noticed that the classification accuracy is usually higher when using data from the transient parts of the signal (rise and decay) than when steady state data is employed, as denoted by the accuracy peaks found around seconds 30 and 90.
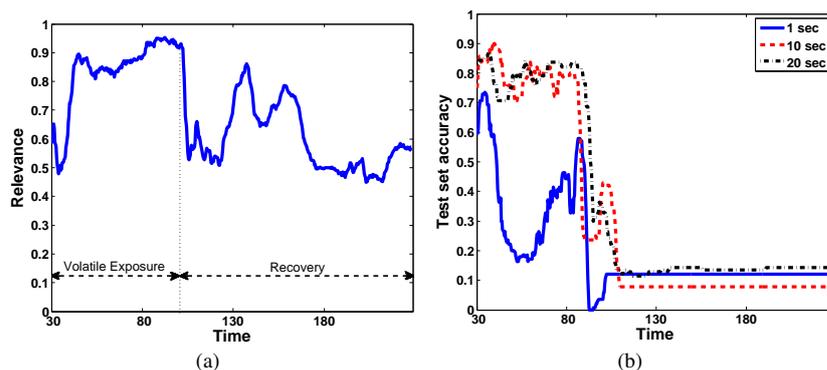
Fig. 7: Time points relevance profile 7(a) averaged over all classes and mean prediction accuracy over time with window length's of $\approx 1, 10, 20sec$ 7(b).

Fig. 8: (a) The robotic arm used in the third experiment mounted over a mobile platform, and a detailed view of the attached e-nose aspiration. (b) Picture of the proposed setup for the third experiment. Each of the black plastic vessels contains a different substance.

5.3 Experiment 3: Robotics experiment - uncontrolled gas pulses

Finally, and with the aim to validate the classification performance in a more challenging robotic scenario, a third experiment is presented. In this case, the e-nose aspiration (see Figure 8) is attached to the hand of a robotic arm [34] which is commanded to approximate the e-nose aspiration to each of four recipients containing different substances (Acetone, Ethanol, Butane [5] and Gin).

To avoid waiting for the sensors to recover their baseline levels after each exposure (which would take more than a minute), we have employed a specially designed e-nose, called MCE-nose [17], that allows the measurement of fast changing gas concentrations.

The robotic arm is commanded to approximate to the containers following a pre-defined sequence. The exposition to each of the substances takes $20sec$, after which the arm moves to another container. The volatile sequence and the gathered signals during the experiment are depicted in Figure 9. A video of a similar experiment is additionally available at `http://mapir.isa.uma.es/mapirwebsite/ index.php/2008-tep-4016-media`

Each of the short sequences was pre-processed such that the baseline is removed. Then the sequences have been matched with the SGMT-TT or NN model as obtained from DS1 [6]. This can be considered to be a test of the model on an independently measured hold out dataset.

The ground-truth and predicted labels of the sequences are given in Table 2 with only 3 errors out of the 16 test samples. In the experiment the SGTM-TT classifier was continuously online and fed by new data every 20sec. according to the measurement protocol. This experiment is interesting because the input data processed by the SGTM-TT method are substantially shorter than the training dataset, with around 30 sampling points for the core measurement. The SVM model can not be applied here due to the varying length of the input data and for the RTK model the sequence are also too short to get reliable predictions as the method is not designed for this type of test inputs. For NN we applied a local DTW alignment between each training and test sample using the best local fit.

6 Conclusion

A novel approach for the analysis of high dimensional and rather short temporal sequences was presented. It is based on the idea to introduce available meta information into the modeling process of a Generative Topographic Mapping through time, given in form of supervised information and relevance learning. We have analyzed the suitability of such model for the odor classification problem in robotics applications, providing comparative results with support vector machine (SVM), nearest neighbor (NN) and the reservoir time series kernel (RTK) for three different scenarios (with

---

[5] Since butane is found at gas state at ambient temperature, the content of a lighter was released when the e-nose aspiration moved over the container.

[6] Here we simply used the model from the first crossvalidation run.
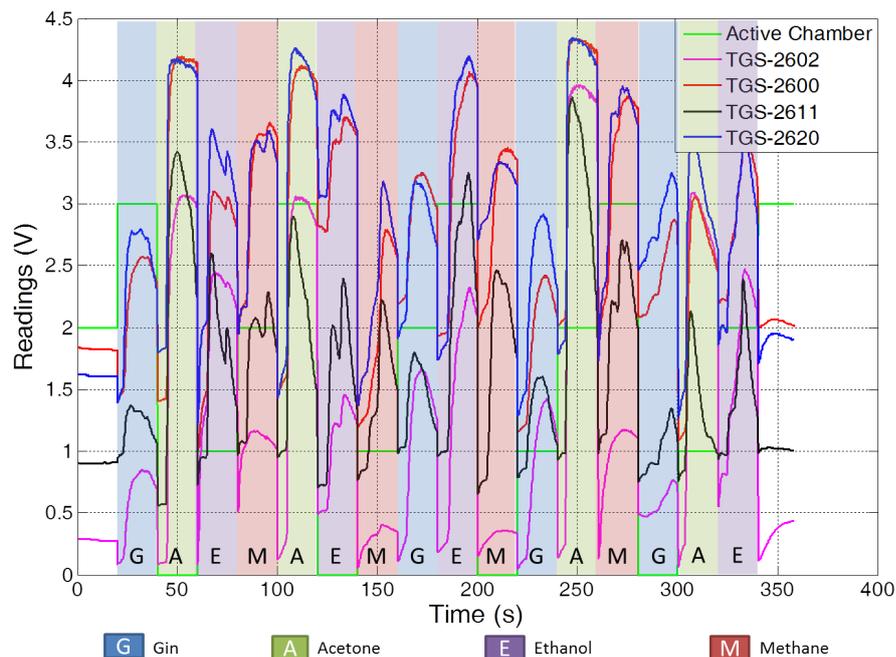
Fig. 9: MCE-nose gathered signals of the classification experiment with a robotic arm, and the "ground-truth" sequence of the employed analytes. The active chamber [0,1,2,3] is switched every 20 sec. Signals are shown for the 4 different sensor channels as described before.

Time	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340
True	G	A	E	M	A	E	M	G	E	M	G	A	M	G	A	E
SGTM-TT																
Pred.	G	M	E	M	A	M	M	G	E	M	G	A	M	G	A	M
Error		o				o										o
NN																
Pred.	G	E	G	M	E	G	G	G	G	M	G	A	G	M	E	G
Error		o	o		o	o	o		o				o	o	o	o

Table 2: Predictions for the external evaluation data using the first respective cross-validation model. The 'o' in the line labeled with *Error* indicates mismatches.

522 increasing classification challenge), and demonstrating that the proposed method is
523 effective for solving such highly dimensional data problem.

524   Other remarkable advantages of the method in the context of odor classification
525 in robotics are: on the one hand, the possibility for the robot to perform rapid classifi-
526 cation of chemical substances by using a short data sequence. On the other hand, the
527 SGTM-TT method outputs relevance values for both the sensors being used as well

as the time-points of the signal, which provide very valuable information to configure the e-nose and to carry out the robot smelling.

In future work it will be of interest to analyze the SGTM-TT in the context of drift problems as recently discussed in [47, 46] and how the method can be further improved by early decision strategies [20].

Acknowledgment

The first author likes to thank: Tien-ho Lin, Carnegie Mellon University, USA for support with the simulation data. Further, Ivan Olier, University of Manchaster, UK; Iain Strachan, AEA Technology, Harwell, UK and Markus Svensen, Microsoft Research, Cambridge, UK for providing code for GTM and GTM-TT. Further we would like to thank Fengzhen Tang, University of Birmingham for providing invaluable support with the RTK method.

References

1. C.A. Astudillo and B.J. Oommen. Topology-oriented self-organizing maps: a survey. *Pattern Analysis and Applications*, pages 1–26, 2014.
2. Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams. Gtm: The generative topographic mapping. *Neural Computation*, 10(1):215–234, 1998.
3. C.M. Bishop. *Pattern recognition and machine learning.* Information science and statistics. Springer, 2006.
4. José-Luis Blanco, Javier G. Monroy, Javier González-Jiménez, and Achim Lilienthal. A kalman filter based approach to probabilistic gas distribution mapping. In *28th Symposium On Applied Computing (SAC)*, mar 2013.
5. Sofiane Brahim-Belhouari and Amine Bermak. Gas identification using density models. *Pattern Recognition Letters*, 26(6):699 – 706, 2005.
6. H. Chen, F. Tang, P. Tino, and X. Yao. Model-based kernel for efficient time series analysis. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (KDD'13)*, Chicago, USA, 2013.
7. Ivan G. Costa, Alexander Schönhuth, Christoph Hafemeister, and Alexander Schliep. Constrained mixture estimation for analysis and robust classification of clinical time series. *Bioinformatics*, 25(12), 2009.
8. C. Distante, P. Siciliano, and K.C. Persaud. Dynamic cluster recognition with multiple self-organising maps. *Pattern Analysis and Applications*, 5(3):306–315, 2002.
9. Cosimo Distante, Marco Leo, Pietro Siciliano, and Krishna C. Persaud. On the study of feature extraction methods for an electronic nose. *Sensors and Actuators B: Chemical*, 87(2):274 – 288, 2002.
10. Ahlame Douzal-Chouakria and C. Amblard. Classification trees for time series. *Pattern Recognition*, 45(3):1076 – 1091, 2012.

11. D. Dumitrescu, B. Lazzerini, and F. Marcelloni. A fuzzy hierarchical classification system for olfactory signals. *Pattern Analysis and Applications*, 3(4):325–334, 2000.

12. Bernd Ehret, Konstantin Safenreiter, Frank Lorenz, and Joachim Biermann. A new feature extraction method for odour classification. *Sensors and Actuators B: Chemical*, 158(1):75 – 88, 2011.

13. Gabriele Ferri, Emanuele Caselli, Virgilio Mattoli, Alessio Mondini, Barbara Mazzolai, and Paolo Dario. Spiral: A novel biologically-inspired algorithm for gas/odor source localization in an indoor environment with no strong airflow. *Robotics and Autonomous Systems*, 57(4):393 – 402, 2009.

14. Jordi Fonollosa, Irene Rodrguez-Lujn, Marco Trincavelli, Alexander Vergara, and Ramn Huerta. Chemical discrimination in turbulent gas mixtures with mox sensors validated by gas chromatography-mass spectrometry. *Sensors*, 14(10):19336–19353, 2014.

15. A. Gisbrecht and B. Hammer. Relevance learning in generative topographic mapping. *Neurocomputing*, 74(9):1359–1371, 2011.

16. Javier González-Jiménez, Javier G.Monroy, and José-Luis Blanco. Robots that can smell: motivation and problems. In *Submitted to 15th International Symposium On Olfaction and Electronic Nose (ISOEN)*, 2013.

17. Javier Gonzalez-Jimenez, Javier G. Monroy, and Jose Luis Blanco. The multi-chamber electronic nose. an improved olfaction sensor for mobile robotics. *Sensors*, 11(6):6145–6164, 2011.

18. R. Gutierrez-Osuna. Pattern analysis for machine olfaction: a review. *Sensors Journal, IEEE*, 2(3):189 – 202, jun 2002.

19. B. Hammer and Th. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.

20. Nima Hatami and Camelia Chira. Classifiers with a reject option for early time-series classification. In *Proceedings of the IEEE Symposium on Computational Intelligence and Ensemble Learning, CIEL 2013, IEEE Symposium Series on Computational Intelligence (SSCI), 16-19 April 2013, Singapore*, pages 9–16. IEEE, 2013.

21. J. Lee and M. Verleysen. Generalizations of the lp norm for time series and its application to self-organizing maps. In Marie Cottrell, editor, *5th Workshop on Self-Organizing Maps*, volume 1, pages 733–740, 2005.

22. A.J. Lilienthal, M. Reggente, M. Trincavelli, J.-L. Blanco, and J. Gonzalez. A statistical approach to gas distribution modelling with mobile robots - the kernel dm+v algorithm. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 570 –576, 2009.

23. Tien-ho Lin, Naftali Kaminski, and Ziv Bar-Joseph. Alignment and classification of time series gene expression in clinical studies. In *ISMB*, pages 147–155, 2008.

24. Paulo J. G. Lisboa. Interpretability in machine learning - principles and practice. In Francesco Masulli, Gabriella Pasi, and Ronald R. Yager, editors, *Fuzzy Logic and Applications - 10th International Workshop, WILF 2013, Genoa, Italy, November 19-22, 2013. Proceedings*, volume 8256 of *Lecture Notes in Computer Science*, pages 15–21. Springer, 2013.

25. Amy Loutfi, Silvia Coradeschi, Achim J. Lilienthal, and Javier Gonzalez. Gas distribution mapping of multiple odour sources using a mobile robot. *Robotica*, 27(2):311–319, June 4 2009.

26. L. Marques, U. Nunes, and A. T. de Almeida. Olfaction-based mobile robot navigation. *Thin Solid Films*, 418(1):51 – 58, 2002. Proceedings from the International School on Gas Sensors in conjunction with the 3rd European School of the NOSE Network.

27. Eugenio Martinelli, Gabriele Magna, Alexander Vergara, and Corrado Di Natale. Cooperative classifiers for reconfigurable sensor arrays. *Sensors and Actuators B: Chemical*, 199(0):83 – 92, 2014.

28. M. Michalak. Adaptive kernel approach to the time series prediction. *Pattern Analysis and Applications*, 14(3):283–293, 2011.

29. Javier G. Monroy, Javier Gonzlez-Jimnez, and Jose Luis Blanco. Overcoming the slow recovery of mox gas sensors through a system modeling approach. *Sensors*, 12(10):13664–13680, 2012.

30. P. Moseley and B. Tofield, editors. *Solid state gas sensors*. Bristol, UK: Adam Hilger, 1987.

31. Iván Olier and Alfredo Vellido. Advances in clustering and visualization of time series using gtm through time. *Neural Networks*, 21(7):904–913, 2008.

32. F. Petitjean, A. Ketterlin, and P. Ganarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011.

33. O.J. Prieto, C.J. Alonso-Gonzlez, and J.J. Rodrguez. Stacking for multivariate time series classification. *Pattern Analysis and Applications*, pages 1–16, 2013.

34. Kinova Robotics. The jaco research edition robotic arm. http://www.kinovarobotics.com/.

35. A.C. Sauve and T.P. Speed. Normalization, baseline correction and alignment of high-throughput mass spectrometry data. In *Proc. Gensips*, 2004.

630 36. F.-M. Schleif, A. Gisbrecht, and B. Hammer. Relevance learning for short high-dimensional time
631   series in the life sciences. In *Proceedings of IJCNN 2012*, pages 2069–2076, 2012.
632 37. F.-M. Schleif, B. Mokbel, A. Gisbrecht, L. Theunissen, V. Dürr, and B. Hammer. Learning relevant
633   time points for time-series data in the life sciences. In *Proceedings of ICANN 2012*, pages 531–539,
634   2012.
635 38. F.-M. Schleif, F.-M. Ongyerth, and T. Villmann. Supervised data analysis and reliability estimation
636   for spectral data. *NeuroComputing*, 72(16-18):3590–3601, 2009.
637 39. P. Schneider, M. Biehl, and B. Hammer. Distance learning in discriminative vector quantization.
638   *Neural Computation*, 21:2942–2969, 2009.
639 40. Ronald E. Shaffer, Susan L. Rose-Pehrsson, and R.Andrew McGill. A comparison study of chemical
640   sensor array pattern recognition algorithms. *Analytica Chimica Acta*, 384(3):305 – 317, 1999.
641 41. R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*. Springer Texts In Statis-
642   tics. Springer, 2000.
643 42. I. G. D. Strachan. *Latent Variable Methods for Visualization Through Time*. PhD thesis, University of
644   Edinburgh, Edinburgh, UK, 2002.
645 43. M. Strickert and B. Hammer. Merge SOM for temporal data. *Neurocomputing*, 64:39–72, 2005.
646 44. Andrzej Szczurek, Bartosz Krawczyk, and Monika Maciejewska. {VOCs} classification based on the
647   committee of classifiers coupled with single sensor signals. *Chemometrics and Intelligent Laboratory
648   Systems*, 125(0):1 – 10, 2013.
649 45. Marco Trincavelli, Silvia Coradeschi, and Amy Loutfi. Odour classification system for continuous
650   monitoring applications. *Sensors and Actuators B: Chemical*, 58:265 – 273, 2009.
651 46. Alexander Vergara, Jordi Fonollosa, Jonas Mahiques, Marco Trincavelli, Nikolai Rulkov, and Ramn
652   Huerta. On the performance of gas sensor arrays in open sampling systems using inhibitory support
653   vector machines. *Sensors and Actuators B: Chemical*, 185(0):462 – 477, 2013.
654 47. Alexander Vergara, Shankar Vembu, Tuba Ayhan, Margaret A. Ryan, Margie L. Homer, and Ramon
655   Huerta. Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators
656   B: Chemical*, 166-167(0):320 – 329, 2012.
657 48. Lloyd R. Welch. Hidden Markov Models and the Baum-Welch Algorithm. *IEEE Information Theory
658   Society Newsletter*, 53(4), December 2003.
659 49. Andrey Ziyatdinov, Jordi Fonollosa, Luis Fernndez, Agustn Gutierrez-Glvez, Santiago Marco, and
660   Alexandre Perera. Bioinspired early detection through gas flow modulation in chemo-sensory sys-
661   tems. *Sensors and Actuators B: Chemical*, 206(0):538 – 547, 2015.