Exploiting Evolution for an Adaptive Drift-Robust Classifier in Chemical Sensing

Stefano Di Carlo[1], Matteo Falasconi[2], Ernesto Sánchez[1], Alberto Scionti[1], Giovanni Squillero[1], and Alberto Tonda[1]

[1] Politecnico di Torino – Torino, Italy
[2] Università di Brescia & SENSOR CNR-INFM– Brescia, Italy

Abstract. Gas chemical sensors are strongly affected by drift, i.e., changes in sensors' response with time, that may turn statistical models commonly used for classification completely useless after a period of time. This paper presents a new classifier that embeds an adaptive stage able to reduce drift effects. The proposed system exploits a state-of-the-art evolutionary strategy to iteratively tweak the coefficients of a linear transformation able to transparently transform raw measures in order to mitigate the negative effects of the drift. The system operates continuously. The optimal correction strategy is learnt without a-priori models or other hypothesis on the behavior of physical-chemical sensors. Experimental results demonstrate the efficacy of the approach on a real problem.

Keywords: real-valued function optimization, parameter optimization, real-world application, chemical sensors, and artificial olfaction.

1 Introduction

Chemical sensing is an intrinsically challenging task. Artificial olfaction [1], also known as *electronic nose*, tries to mimic human olfaction by using arrays of gas chemical sensors together with pattern recognition (PARC) techniques. The most common class of sensors used for chemical sensing is metal oxide semiconductors (MOX) [2]. When sensors come in contact with volatile compounds, the adsorption of these elements on the sensor's surface causes a physical change of the sensor itself, and hence, a change of its electrical properties. Each sensor is potentially sensitive to all volatile molecules in a specific way. The response of the sensor is recorded by its electronic interface, and the corresponding electrical signal is then converted into a digital value. Recorded data are finally elaborated and classified by PARC algorithms (a brief overview of which is presented in Section 2.1) often based on statistical models [3]. The power and appropriateness to data of the PARC strategy determine the final performance of the electronic nose. Yet, independently on that, chemical sensors drift can ultimately invalidate some of the classification models.

Sensor drift is defined as the temporal shift of sensors' response under constant environmental (physical and chemical) conditions. Sensor drift is one of the most serious impairments afflicting all kinds of sensors such as for instance pressure sensors [4], pH sensors [5], conductivity sensors [6], as well as chemical sensors [7]. Sensor drift originates from unknown dynamic processes in the sensor device typically related to sensing

material modifications. These modifications are usually caused by irreversible phenomena such as poisoning or aging. Nowadays, the only effective counteraction to prevent negative effects of drift is frequent sensor calibration. However, while this approach is rather simple to implement for physical sensors where the quantity to be measured is exactly known, chemical sensors pose a series of challenging problems. Indeed, in chemical sensing, the choice of the calibrant strongly depends on the specific application and, when the sensing device is composed of a number of cross-correlated sensors, a univariate signal correction is not feasible.

Long-term drift produces dispersion in the patterns that may change the clusters distribution in the data space. As a consequence, learnt classification boundaries may turn completely useless after a given period of time. Methods for algorithmically correcting, retraining or physically minimizing sensor drift are therefore highly requested by any robust chemical sensing system.

Drift correction algorithms are not new in the field [1] (chapter 13). Notwithstanding the first attempts to tackle this problem date back to early 90s, the study of drift is still a challenging task for the chemical sensor community (see Section 2.2).

In this paper we propose a novel architecture for an evolutionary adaptive drift-robust classifier (Section 3). A linear transformation is applied to the raw measures read by the electronic nose. Such linear transformation is initially described by the identity matrix, and slowly evolved to compensate drift effects in the classification process. The evolutionary core is continuously active, monitoring the performance of the classification and adjusting the matrix coefficients on each new measure. It makes use of a covariance matrix adaptation evolution strategy (CMA-ES), perfectly suited for solving difficult optimization problems in continuous domain. The output of the classifier is used to calculate the fitness function. Compared to existing adaptive solutions, the proposed approach is able to transparently adapt to changes in the sensors' responses even when the number of available samples is not high and new classes of elements are introduced in the classification process at different time frames. To prove this we tested our approach on an experimental data set for detection and classification of chemical compounds by a gas sensor array (Section 4).

2 Background

2.1 Pattern Classification

Pattern recognition is a widely addressed topic in the literature [8]. Automatic recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, remote sensing, and artificial olfaction as well [3].

The primary goal of pattern recognition is supervised or unsupervised classification. In supervised classification the dataset is normally composed of two sub-sets of samples: an initial set called *training set* made of known samples used to train the classifier, and a second group called *test set* composed of unknown samples to classify.

After the signal pre-processing stage, the first critical step in the classification process is feature extraction and/or selection. This phase involves several steps designed to organize and reduce the data space dimensionality, and to avoid problems associated with high-dimensional sparse datasets (course of dimensionality). During the training phase, feature extraction/selection algorithms find the appropriate features to represent the input patterns, and the classifier is trained to partition this resulting feature space.

Dimensionality reduction techniques are also employed for data visualization in order to have a preliminary insight of the multidimensional pattern distribution. Techniques for visualizing and projecting multidimensional data, along with cluster analysis methods, are also referred to as *exploratory data analysis*. The most used exploratory analysis method is principal component analysis (PCA). PCA takes linear combinations of initial variables to identify the directions of maximum variance of the data (called principal components). Typically, only the first two or three components - exploring the highest variance - are retained for visualization purposes, but this is generally enough to understand how data are clustered, i.e., the position and shape of clusters.

Among the various frameworks in which pattern recognition has been traditionally formulated, the statistical approach has been most intensively studied and used in practice [9]. Under this framework several classification algorithms have been proposed and extensively used for chemical sensing, such as: linear discriminant analysis (LDA), k–nearest neighbors (KNN) [10], and more recently support vector machines (SVM) [11] and random forests [12]. Neural networks (NNET) [13] and methods imported from statistical learning theory also received special attention in the field of artificial olfaction [1] (chapter 6).

2.2   Drift Compensation Approaches for Chemical Sensing

Several methods have been proposed to tackle the problem of drift compensation for chemical sensing [1] (chapter 13). Current solutions fall into three categories: (a) use of calibrants to return the classifier to its original state; (b) attune the classifier with proper feature selection/extraction to reduce drift effects; (c) use of "adaptive" models to real-time update the classifier.

Use of a single calibrant or a set of calibrants to retrain a classifier is perhaps the only robust method for determining precise information regarding the degradation of the classification model regardless of the degree of sensor drift [14]. It is also the only method able to sustain a high degree of classification performance even in presence of inconsistent sensor drift over an extremely long period of time. Nevertheless, calibration is the most time-intensive method for drift correction since it requires system retraining. Hence, it should be used sparingly. Moreover, the calibrant selection must be accurately chosen depending on the application. This leads to loss of generalization and lack of standardization, which would be highly required by industrial systems.

Attempts to attune the classifier to specific features of interest have been used in conjunction with both PCA [15] and independent components analysis (ICA) [16] to determine which dimensions of the analyte space most highly correlate with the differences between the analytes in the set. These presumably represent the dimensions that are least noisy and/or are least affected by drift and therefore are the only ones retained in constructing the classification model. Attuning methods can

provide significant improvements in classification over a fixed time period. However, adding new analytes to the recognition library represents a major problem since the previously rejected dimensions might be necessary to robustly identify these new classes. Additionally, these methods contain no provisions for updating the model, and thus may ultimately be invalidated by time evolving drift effects.

Adaptive models try to on-line adjust the classifier by taking into account pattern changes due to drift effects. Neural networks, such as self-organizing maps (SOMs) [17] or adaptive resonance theory (ART) networks [18], have been frequently used in the past. Under such schemes, newly recognized data that match the stored analyte fingerprints, i.e., processed measures used as input for the classifier, can be continuously used to retrain the classifier. This technique has the advantage of simplicity because no recalibration is required. Yet, two main weaknesses can be identified. First, a discontinuity in response between two consecutive exposures (regardless of the time interval between the exposures) would immediately invalidate the classification model and would prevent adaptation. Second, a key to obtain reliable results is to set appropriate thresholds for choosing the winning neuron, and this typically requires a high number of training samples owing to the complexity of the network topology.

3   Proposed Architecture

Figure 1 graphically represents the architecture exploited in this paper to implement a drift-robust classification system for chemical sensing. The proposed approach couples a standard classifier with an adaptive mechanism able to compensate drift effects. It is important to highlight that the classifier itself is not modified, and no additional training is required.

The basic idea is to map the vector of raw measures associated to each sample analyzed by the electronic nose (Rm) with a new fingerprint used as input for the classifier (Fp). Fp is obtained by applying a linear transformation represented by a square matrix C to Rm ($Fp = C \times Rm$).

The classifier receives the produced fingerprints and provides as output a partition of the test set in a given number of classes. The performance of the full classification system can be measured in terms of the percentage of fingerprints correctly classified.

The matrix C evolves while the system is operating with the goal of supporting the classifier in compensating the drift effects. At the beginning C is initialized to the identity matrix, hence, no modification is performed. Whenever a new sample is collected and classified, an evolutionary optimizer slightly tweaks the matrix coefficients in order to increase the robustness of the classification.

The evolutionary scheme is backed up by an *adaptation manager* in charge of deciding whether to activate the evolutionary core that updates the matrix C depending on the classification confidence delivered by the classifier for every fingerprint.

The resulting system is potentially able to compensate any possible drift with the reasonable assumption that the drift should be a relatively slow phenomenon compared to the sampling rate. This allows describing the disturbance as a continuous function over a limited number of consecutive measures. Since parameters are adapted seamlessly, the system is able to compensate disturbance up to quite relevant magnitudes. Considering the drift as a slow phenomenon implies that the initial training data can be considered not affected by it.
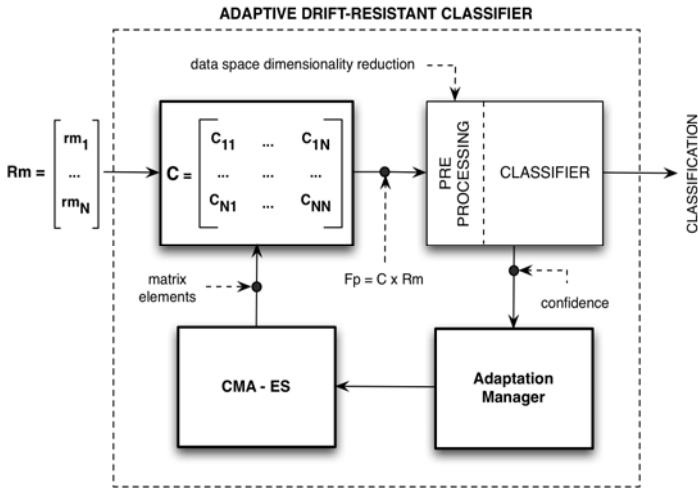
Fig. 1. Main architecture of the adaptive drift-resistant classifier

As evolutionary optimizer we exploit a covariance matrix adaptation evolution strategy. Briefly speaking, an evolution strategy (ES) is a stochastic, population-based, iterative optimization method belonging to the class of evolutionary algorithms, devised in the early 60s by Ingo Rechenberg and Hans-Paul Schwefel. An ES represents an individual as a vector of real-valued numbers. Mutation is performed by adding a normally distributed random value to each vector component. Generating the offspring through mutation corresponds to a sampling in the solution space. The ES is able to determine the optimal value of some of its parameters. Remarkably, the *step size*, i.e., the standard deviation of the normal distribution used for sampling the mutation, is usually self-adapted. The *covariance matrix adaptation* (CMA) is a method to update the covariance matrix of the multivariate normal mutation distribution in the ES. New candidate solutions are generated according to the mutation distribution. Original ES implementations simulate the evolution at the level of species and do not include any recombination operators, although later implementations often do.

The covariance matrix describes the pair-wise relationships between the variables in the distribution. CMA-ES represents the latest breakthrough in the ES field [19]. Results reported in the literature demonstrate that it can easily tackle problems where the fitness landscape presents discontinuities, sharp bends or ridges, noise, and local optima.

In the presented approach, each time a fingerprint is classified with a confidence in a range delimited by an upper-bound and a lower-bound threshold, the adaptation manager enables the CMA-ES. The probability estimates produced by the classifier for the predicted class can be exploited as a measure of the confidence of the classifier in its prediction. This confidence is the fitness value of an individual for the CMA-ES. The two thresholds allow identifying which fingerprints to use to evolve the system, i.e., update the linear transformation *C*, in order to adapt to the drift modifications. The lower threshold aims at discarding spurious fingerprints, i.e., fingerprints whose classification confidence is too low and therefore do not provide reliable

information for the given class, making the proposed framework resistant to spikes. The upper threshold identifies fingerprints that are already corrected in a satisfactory way by the current linear transformation and could not be further enhanced in a sensible way. Skipping these samples allows to reduce the computation effort by activating the evolution only for samples where the effect of the drift becomes more visible.

In order to maximize the correction ability of the proposed architecture, the linear transformation is computed considering the selected fingerprint and a group of K previous fingerprints. The current matrix C is applied to the current measure and to the K previous raw samples. Thus, the evolution continues until the classification confidence for the current fingerprint reaches the upper threshold, and the classification confidence of the K previous ones is not decreased more than ε percent when using the current matrix C.

The final linear transformation matrix is obtained as a linear combination between the current linear transformation matrix and the new one obtained by the application of the CMA-ES, as follows: $C_n = C_n \cdot \alpha + C_{n-1} \cdot (1-\alpha)$, where C_n and C_{n-1} are respectively the new and the current linear transformation matrices, and α is a parameter modeling the *inertia* of the system when moving from the old to the new transformation. This limits the possibility of generating linear transformation matrices unable to improve the classification process. Both ε and α are input parameters of the proposed system with values ranging in the interval $[0,1]$.

4   Experimental Results and Discussion

The proposed approach has been validated on a real data set collected at *SENSOR Lab*, an Italian research laboratory specialized in the development of chemical sensor arrays (more information is available at *http://sensor.ing.unibs.it/*).

The data under consideration have been obtained using the EOS835 *electronic nose* composed of 6 chemical MOX sensors (further information on sensors and equipment can be found in the review paper [2] and its references). The main goal of the performed experiments is to determine the capability of the EOS835 to identify five pure organic vapors, namely: ethanol (1), water (2), acetaldehyde (3), acetone (4), ethyl acetate (5). These are common chemical compounds to be detected in real-world applicative scenarios such as food industry, industrial processes, and biomedical field.

Five different measurement sessions were performed during a period of time of about one month. The elapsed time, though not very long, is enough to obtain data affected by a certain amount of drift.

Different classes have been introduced during the different measurement sessions, a common practice in real world experiments: classes 1 and 2 are measured since the beginning, class 3 is first introduced during the second session (one week later), while classes 4 and 5 appear only during the third session (ten days after the beginning of the experiment).

The number of samples contained in the data set is high (545 measurements) if compared to datasets reported in the literature. It must be noticed that performing measurements with arrays of chemical sensors is a time consuming task that limits the amount of samples that can be produced.

There is not a perfect balance among the number of measurements belonging to every class, with a clear predominance of classes 1, 2 and 3 over classes 4 and 5. This perfectly respects real situations, and additionally increases the difficulty of properly compensating the sensor drift.

PCA plot in figure 2 clearly shows the presence of sensor drift in the considered data set, i.e. shift of samples in a direction over the time. Measures of the test set tend to drift toward a direction that is perfectly visible on the first two principal components. This phenomenon leads to an overlapping of the different classes and, consequently, to a loss of classification performance as time goes on.
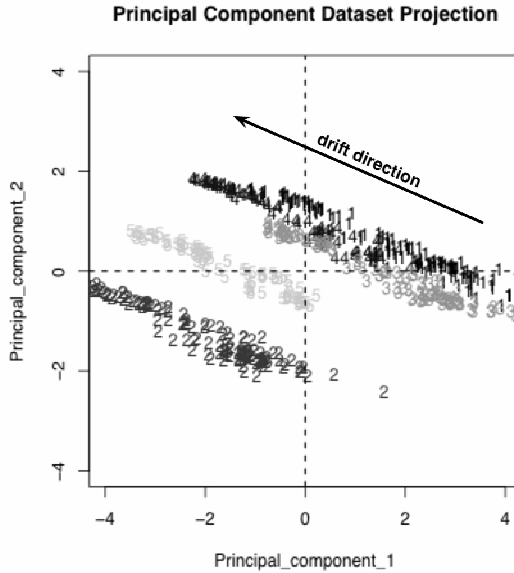

Fig. 2. PCA plot for the test set. The effect of the drift is visible as a shift of the samples over the time towards the upper-left corner.

The classifier used in the experience is a *linear discriminant analysis* (LDA) classification algorithm implemented using the R scripting language [20], while the adaptation block is implemented in C and PERL.

The training set for the classifier is composed of the first 20 measurements of each class, while the remaining 445 samples have been used as test set. The considered training set is quite small compared to common experimental setups that tend to use about one third of the samples to build the prediction model. Nevertheless, this allows to work with a reasonable number of test samples that can be used to better test the evolutionary approach. In order to better understand the characteristics of the considered data model, a preliminary classification of the test set performed with the LDA classifier without drift compensation has been performed. The classifier by itself performs relatively well, with about 90% of the test set correctly classified. Only 43 samples are incorrectly classified. Figure 3-A summarizes this result with a confusion matrix, i.e., a matrix that shows how many samples belonging to a class A have been correctly assigned to class A or misclassified w.r.t. true labels. The figure highlights

how critical samples belonging to class 1, due to the drift, tend to overlap with classes 3 and 4.

The proposed correction system has been applied considering the following parameters, set after 100 experiments aimed at tuning the framework: $\alpha = 0.99$, $\varepsilon = 1$, the lower and upper bounds respectively set to 0.65 and 0.99, and $K = 40$. Small differences in the adopted parameters haven't led to better performance for the proposed system.

The CMA-ES does not require parameter tuning for its application: finding good strategy parameters is considered as part of the algorithm design. In fact, the choice of strategy internal parameters is not left to the user, with the exception of the population size $\lambda = 100$ and the initial standard deviation set to 10^{-2} because of the order of magnitude of the measures in the experiments.

PREDICTION	REFERENCE				
	1	2	3	4	5
1	70	0	0	0	0
2	0	110	0	0	0
3	2	0	92	0	0
4	41	0	0	66	0
5	0	0	0	0	64

(A)

no improvement (10%)
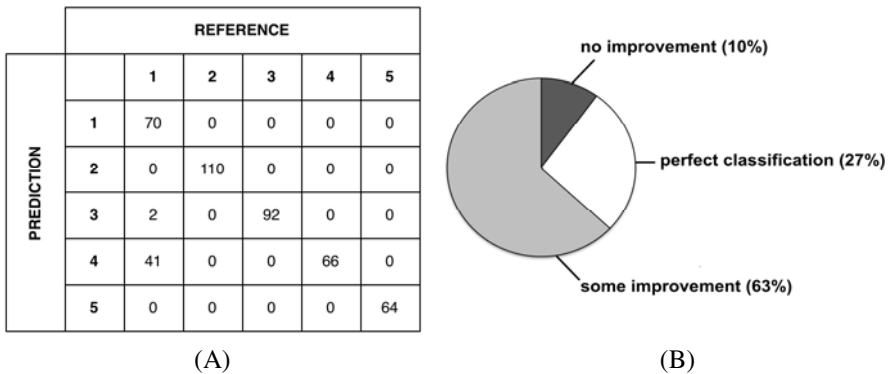perfect classification (27%)
some improvement (63%)

(B)

Fig. 3. (A) Confusion matrix for the not corrected LDA classifier and (B) overall results for the robust classifier

In order to experimentally validate our approach, 150 repetitions of the classification experiments were executed using the previously presented setup (figure 3-B). The proposed approach has improved the original performance of the classifier in 90% of the runs while none of the executions worsened the initial classification. In 27% of the runs all 445 raw measures belonging to the test set have been correctly classified. 63% of the runs produced classification results with an average performance of about 97% approaching the theoretical performance of 100% computed for the LDA algorithm performing cross validation on the training set with 20 iterations and 5% of samples left out at each iteration. The proposed approach can be used for a real-time measurement correction since the time required to obtain a good adjustment (up to 20 mins) is much smaller compared with drift time scale. Moreover it is comparable with the sensors' recovery time.

5   Conclusions

This paper presented an adaptive drift-robust classification system based on an evolutionary strategy. To date, this is the first attempt to exploit evolutionary algorithms capabilities to approach the drift problem in chemical sensing.

The obtained results on a real dataset are encouraging. The classification performances increase to a level close to the theoretical upper bound for the chosen classification algorithm. Yet, while the results are encouraging in terms of improvement of the classification rate, the main drawback of the proposed approach is that it is not able to produce a clear representation of the drift and actually remove it from the row measures. The parameters found by the CMA-ES slightly shift the raw measures in a direction usually orthogonal to the drift, obtaining a better separation of the different classes that allows the LDA classifier to discriminate more accurately among them. This is mainly due to the fact that the evolutionary core of our system does not include any model of the drift that could be exploited to drive the evolution toward a real elimination of the drift effect on the row measures. We are currently working towards the inclusion of several heuristics and considerations on spatial measures in the fitness values, to allow the CMA-ES to both increase the classification rate and perform drift correction. This includes considering different models for the drift effect on chemical sensors.

Comparison between artificially data sets and real data sets will be used to validate our framework.

Acknowledgements

References

[1] Pearce, T.C., Shiffman, S.S., Nagle, H.T., Gardner, J.W.: Handbook of machine olfaction. Wiley-VHC Ed., Weinheim (2003)
[2] Pardo, M., Sberveglieri, G.: Electronic olfactory systems based on metal oxide semiconductor sensor arrays. MRS Bulletin 29(10), 703–708 (2004)
[3] Gutierrez-Osuna, R.: Pattern Analysis for Machine Olfaction: A Review, June 2002, vol. 2, pp. 189–202 (2002)
[4] Polster, A., Fabian, M., Villinger, H.: Effective resolution and drift of Paroscientific pressure sensors derived from long-term seafloor measurements. Geochem. Geophys. Geosyst. 10(Q08008) (2009)
[5] Chen, D.Y., Chan, P.K.: An Intelligent ISFET Sensory System With Temperature and Drift Compensation for Long-Term Monitoring. IEEE Sensor Journal 8(11-12), 1948–1959 (2008)
[6] Owens, W.B., Wong, A.P.S.: An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats by theta–S climatology. Deep-Sea Research Part I-Oceanographic Research Papers 56(3), 450–457 (2009)
[7] Aliwell, S.R., et al.: Ozone sensors based on WO3: a model for sensor drift and a measurement correction method. Measurement Science & Technology 12(6), 684–690 (2001)
[8] Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, Hoboken (2000)
[9] Jain, A.K., Duin, R., Mao, J.: Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 4–37 (2000)

[10] Dasarathy, B.V. (ed.): Nearest neighbor (NN) norms: Nn pattern classification

[11] Pardo, M., Sbreveglieri, G.: Classification of electronic nose data with support vector machines. Sensors and Actuators B: Chemical, 730–737, June 29 (2005)

[12] Pardo, M., Sbreveglieri, G.: Random forests and nearest shrunken centroids for the classification of sensor array data. Sensors And Actuators B-Chemical 131, 93–99 (2008)

[13] Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford (1995)

[14] Sisk, B.C., Lewis, N.S.: Comparison of analytical methods and calibration methods for correction of detector response drift in arrays of carbon black-polymer composite vapor detector. Sensors and Actuators B: Chemical, 249–268, January 24 (2005)

[15] Artursson, T., et al.: Drift correction for gas sensors using multivariate methods. Journal of Chemometrics 14, 711–723 (1999); Special Issue: Proceedings of the SSC6, HiT/TF, Norway (August 1999)

[16] Di Natale, C., Martinelli, E., D'Amico, A.: Counteraction of environmental disturbances of electronic nose data by independent component analysis. Sensors and actuators. B, Chemical 82(2-3), 158–165 (2002)

[17] Marco, S., Ortega, A., Pardo, A., Samitier, J.: Gas Identification with Tin Oxide Sensor Array and Self-Organizing Maps: Adaptive Correction of Sensor Drifts. IEEE Transactions on Instrumentation and Measurement 47, 316–321 (1998)

[18] Vlachos, D.S., Fragoulis, D.K., Avaritsiotis, J.N.: An adaptive neural network topology for degradation compensation of thin film tin oxide gas sensors. Sensors and Actuators B: Chemical, 223–228, December 15 (1997)

[19] Hansen, N.: The CMA evolution strategy: a comparing review, towards a new evolutionary computation. In: Lozano, J.A., Larranaga, P., Inza, I., Bengoetxea, E. (eds.) Advances on estimation of distribution algorithms, pp. 75–102. Springer, Heidelberg (2006)

[20] Kuhn, K.: Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28(5), 1–26 (2008)