Contents lists available at SciVerse ScienceDirect

Sensors and Actuators B: Chemical

Chemical gas sensor drift compensation using classifier ensembles

Alexander Vergara [a,*,1], Shankar Vembu [a,1], Tuba Ayhan [b,2], Margaret A. Ryan [c], Margie L. Homer [c], Ramón Huerta [a]

[a] BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093, USA
[b] Department of Electronics and Communication Engineering, Technical University of Istanbul, Maslak, TR-34469 Istanbul, Turkey
[c] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

ARTICLE INFO

ABSTRACT

Sensor drift remains to be the most challenging problem in chemical sensing. To address this problem we have collected an extensive dataset for six different volatile organic compounds over a period of three years under tightly controlled operating conditions using an array of 16 metal-oxide gas sensors. The recordings were made using the same sensor array and a robust gas delivery system. To the best of our knowledge, this is one of the most comprehensive datasets available for the design and development of drift compensation methods, which is freely reachable on-line. We introduced a machine learning approach, namely an ensemble of classifiers, to solve a gas discrimination problem over extended periods of time with high accuracy rates. Experiments clearly indicate the presence of drift in the sensors during the period of three years and that it degrades the performance of the classifiers. Our proposed ensemble method based on support vector machines uses a weighted combination of classifiers trained at different points of time. As our experimental results illustrate, the ensemble of classifiers is able to cope well with sensor drift and performs better than the baseline competing methods.

1. Introduction

In the history of sensors' development, the electronic noses, or simply *e-noses*, are a relatively new addition to the world of sensors, and can be defined as a collection of broadly cross-reactive sensors connected to electronics and an effective pattern recognition system used to detect, classify, and, where necessary, quantify a variety of chemical analytes or odors of concern in a certain area [1,2]. Ideally, these systems would greatly benefit from chemical sensors that would always show an identical response when exposed to the same analyte or a chemical mixture and return to their baseline level immediately after the gas being evaluated is no longer present at the sensor surface. However, in real-life applications, where sensors are operated over a long period of time, such an ideal situation is still largely unrealizable today [3]. The gradual and unpredictable variation of the chemo-sensory signal responses when exposed to the same analyte under identical conditions, a.k.a. sensor drift, has long been recognized as one of the most serious impairments faced by chemical sensors [4–7].

Drift has plagued the sensor research community for many years, deteriorating the performance of classifiers used for gas recognition and augmenting the maintenance costs of chemo-sensory systems, or artificial electronic noses, during real-time operations. In general, sensor drift can be attributed to two predominant sources [8,9]. First, the 'real-drift' (a.k.a. first-order drift) due to the chemical and physical interaction processes of the chemical analytes, in gas phase, occurring at the sensing film microstructure, such as aging (e.g. the reorganization of the sensor surface over long periods of time) and poisoning (e.g., irreversible binding due to external contamination). And second, the 'second-order drift' (or measurement system drift, among many other names), produced by the external and uncontrollable alterations of the experimental operating system, including, but not limited to, changes in the environment (e.g., temperature and humidity variations); measurement delivery system noise (e.g., tubes condensation, sample conditioning, etc.); and thermal and memory effects (e.g., hysteresis or remnants of previous gases). In general, a number of approaches under the notion of sensor drift counteraction have been implemented in the literature, but one of the pioneering works, and perhaps the most systematic sensor drift analysis was performed by Romain and co-workers [10,11], who utilized a very comprehensive dataset, collected over long periods of time in real operating conditions, to provide a deep insight into the sensor drift problem, for both the real and second-order drift. Among the many interesting conclusions drawn from that work,

the following three aspects were emphasized: (i) from all the sensing technologies available, metal oxide based gas sensors [12] remain the best option for long term applications for continuous monitoring systems; (ii) a calibration gas is recommended to estimate sensor drift compensation, and (iii), sensor replacement is unavoidable over long periods of time.

In practical applications, it is difficult to empirically differentiate between real drift and second-order drift, if possible at all. Accordingly, it is hard to develop methods to correct different sources of drift because the origin of it cannot be ascertained. Utilizing an effective delivery system allows the chemical sensors to bypass the second-order drift effect, making it possible to exclusively concentrate on the chemical sensors for compensating real drift. In our particular gas delivery system, we can control the second-order drift, too, so we can exclusively address the real drift problem. Thus, in the remainder of this document we use the term *drift* to refer to *real drift*. Concerning real drift reduction, many efforts have been devoted to find the sensor materials that can reversibly interact with the gas so that the detected molecules unbind the sensor material as soon as the gas has been purged out from the sensor surface [13–15]. Other solutions based on periodically changing the sensor working temperature [16,17] have also been implemented in an effort to minimize the effects of irreversibility in the sensors' responses due to poisoning. Undoubtedly, heightened reversibility in the sensor response is necessary for the effective drift counteraction. However, this general treatment only constitutes one facet of the problem—the so-called short-term drift—a substantial study of the sensor variability over longer periods of time is also necessary.

The most commonly used solutions to cope with sensor drift within the chemical sensing community are univariate and multivariate methods, where drift compensation is performed either on each sensor individually or on the entire sensor array [9,18]. Among the multivariate drift compensation methods, unsupervised component correction techniques are the most popular [4,19,20]. These techniques rely on finding linear transformations that normalize the sensor responses across time so that a classifier can be directly applied to the resulting stationary data. For instance, the component correction method presented by Årtusson et al. [4] applies the following transformation to the measurement/data matrix $X \leftarrow X - (X \cdot c)c^\top$, where c is the principal component vector(s) of the measurements computed using a reference gas that may approximate the drift direction. The main drawback of these techniques is that they assume the drift direction to be linear in the feature space and, therefore, a linear transformation of the data suffices to correct it. While it is entirely plausible that kernelized versions of component analysis, such as kernel principal component analysis [21], can be applied to account for non-linearities in the drift direction, these techniques have not been investigated much in the chemical sensing community. Also, with the exception of Ref. [20], these techniques require a reference gas that is used to approximate the drift direction by assuming that the reference gas provides the drift direction in all the other gases.

In this paper, we take a completely different approach to solve the mentioned problem, in which we do not make any of the above-mentioned assumptions. Instead, we use a supervised machine learning method, namely, an ensemble of classifiers to cope with sensor drift. To the best of our knowledge, such a machine learning approach, that automatically detects and copes with sensor drift, has not been applied in the chemical sensing community before, although it has been shown to yield promising results on problems with drifting concepts in machine learning and data mining [22–25]. Utilizing a comprehensive dataset of a multi-component gas classification problem recorded from metal-oxide gas sensors over a course of 36 months, we investigate the feasibility of our ensemble of classifiers methodology to mitigate the drift effect in chemical gas sensors. It is important to note that the ensemble method used in this paper *complements*, rather than *competes* against, the existent component correction methods mentioned above; since component correction methods are essentially a pre-processing technique, we can indeed use the ensemble method on the pre-processed data, too. In the remainder of this paper, we first describe the experimental setup, the dataset, and the feature extraction methods considered in this work (Section 2). We then describe the drift compensation algorithm (Section 3), followed by a detailed description of our experimental findings (Section 4). And finally, we present the concluding comments drawn from the results presented in this paper (Section 5).

2. Data collection

We apply our drift compensation method to an extensive dataset[3] recorded by a metal-oxide gas sensor array. In this section, we describe the experimental setup, the recording protocol, and the signal processing algorithms used for feature extraction.

2.1. The experimental setup

We used a sixteen screen-printed commercially available metal-oxide semiconductor gas sensors array, manufactured and commercialized by Figaro Inc. [12], for our experiments. The custom design used in the sensing technology includes an independently controlled RuO_2 (Ruthenium Oxide) electrical heating line and a metal oxide semiconductor film as a sensor material printed onto the measuring electrodes (noble metal). The obtained sensor element is mounted onto an alumina substrate and then connected by lead wires to the pins of the sensor package. The resulting array, populated by sensor devices (4 of each) tagged by the manufacturer as TGS2600, TGS2602, TGS2610, TGS2620 is placed into a 60 ml-volume test chamber, where the odorants of interest, in gaseous form, are to be injected for trials. To generate the required dataset, we connect the said test chamber to a computer-controlled continuous flow system, which provides versatility for conveying the chemical compounds of interest at the desired concentrations to the sensing chamber with high accuracy and in a highly reproducible way while keeping the total flow constant. In particular, our system utilizes three digital, computed-supervised mass flow controllers (MFCs) (provided by Bronkhorst High-Tech B.V. [26]), each of them with different maximum flow levels (200, 100, and 20 ml/min, ±1% of accuracy). Such devices connect to different pressurized gas cylinders, which contain, diluted in dry air, either the carrier gas or the chemical analytes to be measured. To maintain the moisture level constant at 10% R.H. (measured at $25 \pm 1\,^\circ C$) during the entire measurement process, we utilize synthetic dry air as background for all measurements, provided by Airgas Inc. [27]. Then, the analytes under analysis (i.e., ammonia, acetaldehyde, acetone, ethylene, ethanol, and toluene) are added to this background in random order. The total flow rate across the sensing chamber is set to 200 ml/min and kept constant for the whole measurement process. The response of the gas sensor array was measured when the operating temperature of sensors was fixed at $400\,^\circ C$, which, according to the deterministic one-to-one look-up table provided by the manufacturer [12], is attained via a built-in heater that is driven by an external DC voltage source set at 5 V. Finally, to ensure that reproducible response patterns are acquired during each measurement, the sensors were pre-heated for several days prior the experiment process gets started.

---

[3] The dataset will be made available on the UCI repository upon acceptance of the paper.