In this paper, we present a space invariant architecture to enable the Independent Component Analysis (ICA) to solve chemical detection from two unknown mixing chemical sources. The two sets of unknown paired mixture sources are collected via JPL 16-ENose sensor array in the unknown environment with, at most, 12 samples data collected.

Our space invariant architecture along with the maximum entropy information technique by Bell and Sejnowski and natural gradient descent by Amari has demonstrated that it is effective to separate the two mixing unknown chemical sources with unknown mixing levels to the array of two original sources under insufficient sampled data. From separated sources, they can be identified by projecting them on the 11 known chemical sources to find the best match for detection. We also present the results of our simulations. These simulations have shown that 100% correct detection could be achieved under the two cases: a) under-completed case where the number of input (mixtures) is larger than number of original chemical sources; and b) regular case where the number of input is as the same as the number of sources while the time invariant architecture approach may face the obstacles: overcomplete case, insufficient data and cumbersome architecture.

Keywords: space invariant independent component analysis, ENose, chemical detection

1. Introduction

The need for low-power, miniature sensor devices that can monitor air quality in an enclosed space with multi-compound capability and minimum human operation has led to the development of polymer-carbon composite based electronic nose (ENose) at NASA’s Jet Propulsion Laboratory (JPL) [1–3]. The sensor array in the JPL ENose consists of 32 conductometric sensors made from insulating polymer films loaded with carbon. In its current design, it has the capability to detect 10 common contaminants which may be released into the recirculated breathing air of the space shuttle or space station from a spill or a leak; target concentrations are based on the 1-hour Spacecraft Maximum Allowable Concentrations (SMAC) set by NASA (see Table 1) [4], and are in the parts-per-million (ppm) range. The ENose was intended to fill the gap between an alarm which has little or no ability to distinguish among chemical compounds causing a response and an analytical instrument which can distinguish all compounds present but with no real-time or continuous event monitoring ability.

As in other array-based sensor devices, the individual sensor films of the ENose are not specific to any one analyte; it is in the use of an array of different sensor films that gases or gas mixtures can be uniquely identified by the pattern of measured response. The response pattern requires software analysis to deconvolute gas compounds and their concentrations.

An example of a sensor set is shown in Fig. 1, and the complete device on which the data was used in this study is shown in Fig. 2.

The specific analysis scenario considered for this development effort was one of leaks or spills of specific compounds. It has been shown in analysis of samples taken from space shuttle flights that, in general, air is kept clean by the air revitalization system and contaminants are present at levels significantly lower than the SMACs [5]; the JPL ENose has been developed to detect target compounds released suddenly into the breathing environment. A leak or a spill of a solvent or other compound would be
an unusual event. Release of mixtures of more than two or three compounds would be still more unusual; such an event would require simultaneous leaks or spills to occur from separate sources. Thus, for this phase of development, mixtures of more than two target compounds were not considered.

In this paper, we consider an approach to analysis of sensor responses to mixtures so that use of the JPL ENose may be extended to detection of chemical compounds in an open and changing environment, such as a building or a geographical area where air exchange is not controlled and limited.

As such operation in the open environment, the collected sensory data will be a mixing between the unknown chemicals with the unknown mixing levels (coefficient) between them. The identification of the chemical compounds among these mixing chemicals is a real challenge for real world applications.

To search for a chemical compound whether it exists in the operating environment, one of the most robust techniques is to recover the original chemicals. When done, the detection can be an easy step by finding the maximum correlation between the predicted original chemicals and the target chemicals. More sophisticated way, the neural network approach [6–10] can be employed to capture the target chemicals in various conditions through learning, e.g., concentration levels through the parameterized weight set, then the strongest correlation between parameterized weight and the predicted original can be used to identify the intended chemical.

Recently, Independent Component Analysis (ICA) [11–19] has proven that it not only de-correlates the second order statistics of the signals but also reduces the higher order statistical dependencies. ICA transforms an observed signal vector into a set of signals that are as statistically independent as possible. Theoretically, ICA is an information-theoretic approach, which exploits concepts from information theory such as entropy and mutual information.

The ICA roots in the early work of Herault and Jutten [11] who first introduced an adaptive algorithm in a simple feedback architecture that was able to separate several unknown independent sources. ICA was further developed by [12–17]. Amari et al. [18] have used natural gradient descent based on the Riemannian metric tensor to optimize the curvature of a particular manifold in n dimensional space. This technique is employed to apply to the Infomax [16] to simplify the learning rule that is used in this paper. ICA has applications for feature extraction in speech recognition systems, in communication systems, in medical signal processing, and in image processing.

2. Technical Approach

In this study, we based on the approximation of the set up architecture shown in Fig. 3. In Fig. 3, the collected sensing data \( x_i(t) \) consists of changes in electrical resistance corresponding to sensor response to the unknown mixture \( \sum_{j=1}^{N} \alpha_j s_j(t) \) of chemical sources \( s_j(t) \) and their densities (or concentrations) \( \alpha_j \) at the time \( t \).

Due to the small spatial separation between the sensors themselves, the input of chemical exposure on each sensor is assumed to be uniquely distributed. The sensory data can be modeled as follows:

\[
  x_i(t) = f_i(\alpha_1 s_1(t), \ldots, \alpha_N s_N(t)) \quad \ldots \ldots \quad (1)
\]

where \( f_i \) is the unknown non linear activation function, \( \alpha_j \)
is the unknown mixing coefficient of chemical source \(j\), and \(i\) is the index of sensor number and \(N\) as number of chemical sources.

Assume that there exists an operating point \(v^o_0 = \{v^o_1, \ldots, v^o_N\}\) and \(v^o_j = \sum_{j=1}^{N} \alpha_j s^o_j\) so that the Eq. (1) can be approximated by the first order of a Taylor expansion and it is written:

\[
x_i = a_i + \sum_{j=1}^{N} \gamma_j s_j \tag{2}
\]

where

\[
a_i = f_i(\alpha_1 s^o_1, \ldots, \alpha_N s^o_N) = \frac{\partial f_i}{\partial v_1} \bigg|_{v^o_j} \sum_{j=1}^{N} \alpha_j s^o_j \tag{3}
\]

and

\[
\gamma_j = \frac{\partial f_i}{\partial v_1} \bigg|_{v^o_j} \alpha_j \tag{4}
\]

and \(s^o_j\) is an operating point of the source \(s_j\).

For each sampling data point in time for the same sensor \(i\), \(x_i\) fluctuates around its bias point \(a_i\) and it can be considered as a common bias for \(x_i(t)\) with \(t \in [t_i + k\Delta t]\). From this argument, Eq. (2) can be simplified to:

\[
Y = \begin{bmatrix}
x_1 - a_1 \\
x_2 - a_2 \\
\vdots \\
x_k - a_k
\end{bmatrix} \tag{5}
\]

\[
Y = \Gamma S \tag{6}
\]

where \(Y\) is unbiased mixture data, \(\Gamma\) is the unknown mixing matrix and \(S\) is the chemical source signal.

The learning rule based on the maximum entropy algorithm [16] is given by:

\[
y = g(u) = g(Wx) \tag{7}
\]

where \(g\) is a non linear function e.g. the logistic function or hyperbolic tangent function. The update weight can be calculated as:

\[
\Delta W = W^{-T} + \Phi(u)x^T \tag{8}
\]

where \(W^{-T}\) as an inverse transport of the NNX weight matrix \(W\), \(x^T\) as a mixing input vector (observed vector), and

\[
\Phi(u) = [\phi_i(u)] = \frac{\partial y_j}{\partial y_i} \tag{9}
\]

To simplify Eq. (7) using a natural gradient descent by Amari [18], the learning rule can be:

\[
\Delta W = (I + \Phi(u)u^T)W \tag{10}
\]

with \(u = Wx\). The details can be found in [19].

3. Space Invariant Architecture

The most common ICA approach is that the number of variables and the number of sources are the same. However, in this study we faced two obstacles: 1) there are 12 or less samples (mixing chemical compounds) from each sensor and the total number of sensor is 16 which it does not have sufficient data set (at least 16 data samples required); and 2) the number of variables is 16 as number of sensors while the number of compounds in a mixture is 2 and it is considered over complete case.

For the time invariant approach, the data that will flow perpendicularly (dotted arrow) to the time invariant direction as shown in Fig. 4 will require 32 outputs (16 channels for each chemical compound). The architecture is 16 inputs, 32 outputs and 12 or less sample data which may not be a solvable approach. Instead of using a time invariant approach, we use the space invariant approach (Fig. 4 with dashed and dotted arrow) which allows us to have more data points and enable the square mapping matrix (the dimension of mixing sources and sensors are the same). This approach is feasible due to the mathematical model based on Eqs. (2) and (6). The architectures are shown in Fig. 4.

In Fig. 4, the unbiased input \(Y_i(t) (i = 1, \ldots, k)\) is based on temporal mixture data and the sensory data are spatially invariant.

From laboratory set up, we have collected a set of single of spectra of 11 chemicals using 16 elements in the ENoSe sensor array; it is averaged and shown in Fig. 5.

In this study, the number of sensors used is 16 and mixing chemical sources is 2; we will examine two sets of data, as shown in Table 2.
Table 2. Parameters of chemical mixture.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Num. of Variables</th>
<th>Mixture</th>
<th>Num. Samples Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>1 and 7</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>3 and 10</td>
<td>9</td>
</tr>
</tbody>
</table>

Fig. 6. The mixture data of chemical 1 and 7.

4. Simulation Results

Based on the data available provided in Table 2, we divided this study into two experiments: 1) Under complete case: in this case we used complete sample data (see column 4 in Table 2) as input to the network and the output size is 2 as two original sources recovered. 2) Squared case: this is a straight forward with 2 inputs and 2 outputs case when the data was rearranged so that each input is from the same sensor with non-overlapped consecutive sampling times: \( t + i\Delta t \) and \( t + (i + 1)\Delta t \).

4.1. Experiment 1

In this case, we have studied the two data sets in Table 2 and we used all data vectors available (The maximum number of data vectors is 12), which is less than the number of sensors (16). For data set 1, there are 12 mixtures of chemical 1 and 7 and the data are shown in Fig. 6.

4.1.1. Data Set 1

Using the space invariant ICA approach, the recovered signal sources (chemical 1 and 7 sources) are shown in Fig. 7 and the average of the single chemical source 1 and 7 are shown in Fig. 8. Figs. 7 and 8, show a strong correlation between the chemical 1 and separated chemical named 1 (* traces) and chemical 7 and separated signal named 7 (+). To confirm its performance, we projected the separated sources 1 and 7 by ICA technique on the known 11 chemical sources shown in Fig. 3, the results are provided in Table 3.

From Table 3, it has demonstrated that the single source of chemical 1 has the greatest overlap (bold face value) with the separated source, labeled separated chemical 1. Similarly, single source chemical 7 has the greatest overlap (bold face value) with separated chemical 7 from the mixture shown in Fig. 6.

Table 3. The Projection of the separated chemical sources 1 and 7 on the original chemical.

<table>
<thead>
<tr>
<th>Sing. chem. source</th>
<th>Separated chemical 1</th>
<th>Separated chemical 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8457</td>
<td>0.9601</td>
</tr>
<tr>
<td>2</td>
<td>0.4772</td>
<td>0.6291</td>
</tr>
<tr>
<td>3</td>
<td>0.7325</td>
<td>0.9369</td>
</tr>
<tr>
<td>4</td>
<td>0.6286</td>
<td>0.5258</td>
</tr>
<tr>
<td>5</td>
<td>0.7433</td>
<td>0.6996</td>
</tr>
<tr>
<td>6</td>
<td>0.4872</td>
<td>0.4304</td>
</tr>
<tr>
<td>7</td>
<td>0.7944</td>
<td><strong>0.9876</strong></td>
</tr>
<tr>
<td>8</td>
<td>0.7938</td>
<td>0.9014</td>
</tr>
<tr>
<td>9</td>
<td>0.6679</td>
<td>0.5737</td>
</tr>
<tr>
<td>10</td>
<td>0.5271</td>
<td>0.6051</td>
</tr>
<tr>
<td>13</td>
<td>0.6223</td>
<td>0.7896</td>
</tr>
</tbody>
</table>
4.1.2. Data Set 2

Data set 2, a mixture of chemicals 3 and 10, is shown in Fig. 9. For the mixture from data set 2, the performance of space invariant ICA has demonstrated its effective capability to separate the mixture of chemical 3 and 10 as shown in Figs. 10 and 11.

In Table 4 it shows the maximum correlation (bold face values) between the original and separated source of chemical 3 and 10.

<table>
<thead>
<tr>
<th>Sing. chem source</th>
<th>Sep. chem 3</th>
<th>Sep. chem 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8250</td>
<td>0.7723</td>
</tr>
<tr>
<td>2</td>
<td>0.4760</td>
<td>0.8282</td>
</tr>
<tr>
<td>3</td>
<td><strong>0.9464</strong></td>
<td>0.8254</td>
</tr>
<tr>
<td>4</td>
<td>0.5235</td>
<td>0.4420</td>
</tr>
<tr>
<td>5</td>
<td>0.6134</td>
<td>0.8935</td>
</tr>
<tr>
<td>6</td>
<td>0.3414</td>
<td>0.7670</td>
</tr>
<tr>
<td>7</td>
<td>0.8681</td>
<td>0.7829</td>
</tr>
<tr>
<td>8</td>
<td>0.8165</td>
<td>0.8887</td>
</tr>
<tr>
<td>9</td>
<td>0.6561</td>
<td>0.7803</td>
</tr>
<tr>
<td>10</td>
<td>0.6125</td>
<td><strong>0.9619</strong></td>
</tr>
<tr>
<td>13</td>
<td>0.7200</td>
<td>0.9444</td>
</tr>
</tbody>
</table>

Table 4. The Projection of the separated chemical sources 3 and 10 on the original chemical.

Table 5. The mean and standard deviation of its projection of separated sources C1 and C7 on each single chemical source.

<table>
<thead>
<tr>
<th>Sing. chem source</th>
<th>Mean Sep. chem. 7</th>
<th>Stand. dev.</th>
<th>Mean Sep. chem. 1</th>
<th>Stand. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8851</td>
<td>0.0830</td>
<td><strong>0.8959</strong></td>
<td>0.0890</td>
</tr>
<tr>
<td>2</td>
<td>0.5893</td>
<td>0.1405</td>
<td>0.5551</td>
<td>0.0792</td>
</tr>
<tr>
<td>3</td>
<td>0.8728</td>
<td>0.0784</td>
<td>0.8046</td>
<td>0.0397</td>
</tr>
<tr>
<td>4</td>
<td>0.4973</td>
<td>0.0723</td>
<td>0.5057</td>
<td>0.1541</td>
</tr>
<tr>
<td>5</td>
<td>0.6611</td>
<td>0.0599</td>
<td>0.7285</td>
<td>0.0742</td>
</tr>
<tr>
<td>6</td>
<td>0.4161</td>
<td>0.0970</td>
<td>0.4465</td>
<td>0.0903</td>
</tr>
<tr>
<td>7</td>
<td><strong>0.9222</strong></td>
<td><strong>0.0765</strong></td>
<td>0.8664</td>
<td>0.1104</td>
</tr>
<tr>
<td>8</td>
<td>0.8398</td>
<td>0.0604</td>
<td>0.8141</td>
<td>0.1099</td>
</tr>
<tr>
<td>9</td>
<td>0.5539</td>
<td>0.0460</td>
<td>0.5601</td>
<td>0.1031</td>
</tr>
<tr>
<td>10</td>
<td>0.5671</td>
<td>0.1103</td>
<td>0.5702</td>
<td>0.1141</td>
</tr>
<tr>
<td>13</td>
<td>0.7342</td>
<td>0.1096</td>
<td>0.6902</td>
<td>0.1101</td>
</tr>
</tbody>
</table>

4.2. Experiment 2

In this experiment, we paired data set 1 (16 × 12) in columns to obtain the data set 96 × 2. From this conversion, data values in a single row are the data from the same sensor with consecutive sampling times t + iΔt and t + (i + 1)Δt; this new data set allows the same number of mixing sources and of original sources. Using this new data set, space invariant ICA has produced the results that were validated with the 11 known chemicals. Classification was 100% correct, based on the projection on 11 classes shown in Fig. 5. To simplify the results, we tabulated the mean and standard deviation of its projection (separated sources C1 and C7) on each single chemical source and the results are summarized Table 5.

Table 5 is compact information to show that the separation sources from the mixture have successfully identified the original chemical sources.

Due to limited data set, the distributions based on the mean and standard deviation in Table 5 are not Gaussian distributions, according to the Central Limit Theorem.
5. Discussions

To separate two (2) mixing sources from a sixteen (16) element sensory data array, known as the over complete case, poses a challenge for mathematical and network architecture. The non overlapped pair-wise (i.e. sensor $i$ and sensor $i+1$) or overlapped pair-wise (i.e. sensor $i$ and sensor $i+1$) and (sensor $i+1$ and sensor $i+2$) so on may face cumbersome and ineffective techniques.

As shown above, the mathematical model has demonstrated space invariant ICA to be an effective architecture to overcome insufficient data samples and the over complete case. Moreover, the chemical data itself is fuzzy and inconsistent, and the optimal architecture is not answered in this study. By simulation we have shown that chemical source separation problem can be solved effectively with complete time sampling data ($k = 12$) (under complete case) and two consecutive sampling data ($k = 2$). Optimal architecture may require a model of noise in order to determine the size of the sampling input. Moreover, space invariant ICA governed by Eq. (6) is only valid when the sampling time is sufficiently small. Hence, the sampling time also plays an important role to ensure that the model approach holds.

6. Conclusions

We have provided a mathematical model to enable the space invariant ICA architecture from which Informax and natural gradient descent technique can be applied and simulation has confirmed that our modeling is effective and sufficient to perform chemical source separation to enable the smart ENoSe.

Further study will be conducted to validate its usefulness for the real world and open environment for chemical detection. In addition, the miniaturized, compact, light weight and low power hardware approach is also a driven force for NASA mission from which System-On-a-Chip approach will be our next focus based on this modeling approach.

Acknowledgements

The research described herein was performed by the Jet Propulsion Laboratory, California Institute of Technology under contract with the National Aeronautics and Space Administration (NASA). The authors would like to thank Drs. Thomas Prince and Katherine Dumas for their support and Dr. T. Daud for his discussion and suggestions. The authors would like to thank the anonymous reviewers for constructive comments and insightful suggestions to make our paper more concrete and readable.

References:


Name: Tuan A. Duong
Affiliation: Jet Propulsion Laboratory/California Institute of Technology (JPL/CIT)

Address: 4800 Oak Grove Dr, Pasadena, CA 91109

Brief Biographical History:
1987: Joined Jet Propulsion Laboratory as Member of Technical Staff
1995-present: Consultant for several high tech companies in US
1995-present: Founder of Adaptive Computation LLC
1998-present: Senior Scientist, JPL/CIT
1999-2003: Adjunct Associate Professor, the EE Dept., Systems, USC

Main Works:
- Friendly Hardware Algorithm Development and Hardware Implementations
- Dynamic Optimization
- Cognitive Computing Architecture
- Real Time Dynamic Target Recognition and Tracking
- Chemical Detection in the heterogeneous environment using ENose, Ion Mobility Spectrometry (IMS), Gas chromatography/Mass Spectrometry (GC/MS), Differential Ion Mobility Spectrometry (DMS) and advanced signal processing approach
- Bio-Inspired Systems

Name: Vu A. Duong
Affiliation: Jet Propulsion Laboratory/California Institute of Technology (JPL/CIT)

Address: 4800 Oak Grove Dr, Pasadena, CA 91109

Brief Biographical History:
1998: Finished BS in Computer Engineering, University of California, San Diego
2000: MS in Electrical and Computer Engineering, University of California, Irvine
2004: Ph.D. in Electrical and Computer Engineering, University of California, Irvine
2004-present: Staff Scientist in Bio-Inspired Technologies and Systems group

Main Works:
- High Order Statistical Analysis
- Evolutionary Computation
- Neural networks
- Intelligent Control Systems
- Evolvable and Reconfigurable Computing Architecture

Name: Margaret A. Ryan
Affiliation: Jet Propulsion Laboratory/California Institute of Technology (JPL/CIT)

Address: 4800 Oak Grove Dr, Pasadena, CA 91109

Brief Biographical History:
Principal Member of Engineering Staff, JPL
Ph.D. Physical Chemistry, University of Massachusetts
Research experience in electrochemistry, sensor materials, sensing systems

Main Works: