Abstract

Locally Linear Embedding as Nonlinear Feature Extraction to Discriminate Liquids With a Cyclic Voltammetric Electronic Tongue

Jersson X. Leon-Medina [1,*], Maribel Anaya [2] and Diego A. Tibaduiza [3]

[1] Departamento de Ingeniería Mecánica y Mecatrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; jxleonm@unal.edu.co
[2] MEM (Modelling-Electronics and Monitoring Research Group), Faculty of Electronics Engineering, Universidad Santo Tomás, Bogotá 110231, Colombia; maribelanaya@usantotomas.edu.co
[3] Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; dtibaduizab@unal.edu.co
* Correspondence: jxleonm@unal.edu.co

Abstract: EElectronic tongues are devices used in the analysis of aqueous matrices for classification or quantification tasks. These systems are composed of several sensors of different materials, a data acquisition unit, and a pattern recognition system. Sensors can be of the electrochemical type for the sensor array systems which are used for the inspection by different methods including cyclic voltammetry. By using this method, each sensor yields a voltammogram that relates the response in current to the change in voltage applied to the working electrode. A great amount of data is obtained in the experimental procedure which allows handling the analysis as a pattern recognition application, however, the development of efficient machine learning based methodologies still an open research interest topic. As a contribution, this work presents a novel data processing methodology is developed for sensor arrays of a cyclic voltammetry electronic tongue. This methodology is composed of several stages such as data normalization through group scaling method and a nonlinear feature extraction step with locally linear embedding (LLE) technique. A reduced-size feature vector is obtained that serves as input to a K-Nearest Neighbors (KNN) supervised classifier algorithm. A Leave-one-out cross-validation procedure is performed to obtain the final classification accuracy. The methodology is validated with a data set of five different juices as liquid substances. Results of this validation show that 80% of classification accuracy was obtained by applying this methodology to this dataset.

Keywords: electronic tongue; locally linear embedding; cyclic voltammetry; K-Nearest Neighbors; classification; machine learning

Citation: Leon-Medina, J.X.; Anaya, M.; Tibaduiza, D.A. Locally Linear Embedding as Nonlinear Feature Extraction to Discriminate Liquids With a Cyclic Voltammetric Electronic Tongue. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published: 1 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Discriminate different types of liquid substance is a daily task in the food industry. This procedure can be used to preserve the flavor of a product, identify adulterations, confirm the presence of a specific liquid, among others [1]. Generally, the analysis of liquid food products is carried out using a panel of previously trained experts[2] who allow tasting and identifying a specific flavor. This through the training of the human sense of taste. However, over time this ability may be deteriorated and, human reliability may be a risk factor for the process. Another method used in the analysis of liquids is high-performance liquid chromatography (HPLC)[3], but this type of analysis is expensive and must be performed in laboratories with specialized equipment. As an alternative to the two mentioned methods, the electronic tongues have emerged because its advantages such as the portability, reliability and low price [4]. Inspired by the human sense of taste and the behavior of taste buds, electronic tongues use an array of non-selective sensors to capture signals from a specific liquid. To do this, these arrays use sensors of different materials and subsequently a sensor data fusion analysis and pattern recognition algorithms to perform classification tasks of different liquids.

In this work, the cyclic voltammetry technique was used to perform analysis experiments on 5 different liquids, two voltametric sensors of the screen-printed electrodes type of two different materials were used in their working electrodes, in this case platinum and graphite. The amount of data captured when performing cyclic voltammetry experiments is high, therefore, these data have high dimensionality. This work uses the Locally Linear Embedding (LLE) [5] method to perform a dimensionality reduction of the original data. This dimensionality reduction serves as feature extraction method that is used as input of a k-Nearest Neighbor (k-NN) [6] classifier used as supervised machine learning method. In order to classify the 5 different juices a Leave One Out cross validation procedure is executed due to the small quantity of samples in the dataset, along to prevent over-fitting [7]. The results show a correct classification procedure of the juices evidenced with a high classification accuracy. The remainder of this papers is as follows: section 2 describes the experimental setup. Next, the section 3 shows the cyclic voltammetry test performed. Following, section 4 presents the data processing results including data unfolding, data scaling, dimensionality reduction, classification, and cross validation. Finally, the section 5 exhibits the main conclusions of the work.

2. Experimental setup for the acquisition of the juice dataset

The methodology developed in this work is used to classify 5 different classes of juices. This dataset of juices was obtained by conducting experiments on 5 different juices from a company located in the city of Tunja in the department of Boyacá-Colombia. Cyclic voltammetry tests were performed on each one of the 5 juices. For each juice, 5 experiment were performed as it is evidenced in Table 1.

Table 1: Description of the type of juice in the dataset registered with the EVAL-AD5940ELCZ potentiostat.

ID	Juice	number of samples
1	*BICHES* FRUITS	5
2	GREEN APPLE	5
3	RED FRUITS	5
4	PASSION FRUIT	5
5	ORANGE	5

Experiments were performed on the different juices using the EVAL-AD5940ELCZ [8] electrochemical evaluation board from Analog Devices. This board is commanded by the evaluation board EVAL-ADICUP3029, which is an Arduino and PMOD compatible development board that includes Bluetooth and WiFi connectivity [9]. The EVAL-ADICUP3029 board uses the ADuCM3029 ultra low power Arm Cortex-M3 processor as the main device. The ADuCM3029 is an integrated mixed-signal microcontroller system for processing, control, and connectivity. The electronic boards EVAL -AD5940ELCZ, EVAL-ADICUP3029 and their integration are shown in Figure 1.

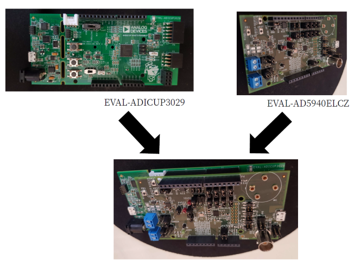

Figure 1. Ensemble of the EVAL-AD5940ELCZ and EVAL-ADICUP3029 boards.

The integration of the EVAL -AD5940ELCZ and EVAL-ADICUP3029 boards is used as potentiostat equipment. This system provide only 1 channel in such a way that a cyclic voltammogram was obtained at a time, In the experimentation the sensor had to be changed to perform each cyclic voltammetry experiment. This electronic tongue used two reference screen printed electrode sensors, AC1.W2.R2 DW = 1 and AC1.W4.R2 DW = 1. The first one was used as a working electrode of platinum and the second one of graphite. AgCl-covered silver was used for the reference electrodes. The experimental configuration used to obtain the data set of 5 juices is depicted in figure 2.

Figure 2. Integrated electronic tongue system. Computer with Sensor Pal software, USB cable to the EVAL-AD5940ELCZ and EVAL-ADICUP3029 attached boards, colored crocodile cable, sensor connection cable and SPE Sensor.

3. Cyclic voltammetry tests to obtain the juice data set

The Sensor Pal command software from Analog Devices was used to perform the cyclic voltammetry tests. The parameters used in the development of these experiments are shown in Table 2. The ramp-type drive signal shown in blue in Figure 3 has a total duration of 4 seconds. The number of the response current signal points is equal to 500 since there is a period of 8ms for each sample corresponding to the scan rate used, which is equal to 500mV/s. Results shown by the green line in the unfolding voltammogram present data current in the ordinate axis in the order of μA .

Table 2: Parameters used in cyclic voltammetry tests to obtain the juice data set.

Parameters	Value	
Initial potential	-1000	mV
Final potential	1000	mV
Potential step	4	µV
Scan rate	500	mV / s
Current range	± 450	µA
Calibration resistance	12000	Ω
Load resistance	100	Ω

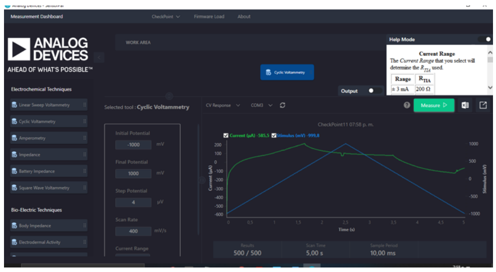

Figure 3. Cyclic voltammetry test in Sensor Pal software from Analog Devices. Blue line represents the excitation ramp signal, while the green line shows the response signal.

Figures 4 and 5 show the cyclic voltammograms for two different juices with both the platinum and graphite sensors in the electronic tongue system by using the boards EVAL -AD5940ELCZ and EVAL-ADICUP3029 as potentiostat. In particular, Figure 4 depicts the cyclic voltammograms obtained for green apple juice showing that the voltammogram obtained by the graphite sensor reaches higher positive current values than the platinum sensor. In contrast, Figure 5 shows the cyclic voltammograms for an experiment in red fruit juice, the magnitude of the current obtained by the graphite sensor is clearly lower than with the platinum sensor.

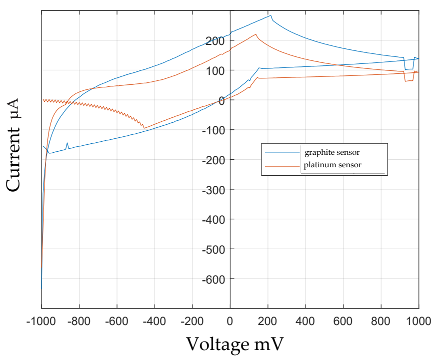

Figure 4. Cyclic voltammograms obtained by platinum and graphite sensors in experiment # 2 on juice # 2 (green apple).

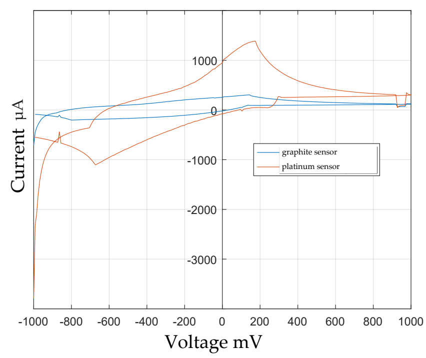

Figure 5. Cyclic voltammograms obtained by platinum and graphite sensors in experiment # 5 on juice # 3 (red fruits).

4. Data processing Results

4.1. Data Unfolding

The unfolding of the cyclic voltammogram data obtained by each sensor is carried out according to the group scaling method [10]. For each experiment carried out, the unfolding of the two sensors is performed, obtaining a signal of 1000 data points. Figure 6 shows an unfolded signal by juice number 3 (red fruits). In this case, the ordinates correspond to current measurements in mA and the abscissa to data points. Since 25 juice samples were considered in total, the matrix size X is equal to 25 $\times$1000.

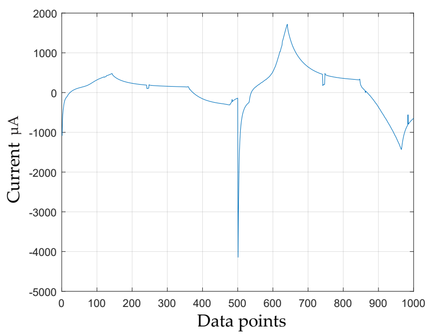

Figure 6. Example of unfolded signal of the cyclic voltammograms obtained by the platinum and graphite sensors.

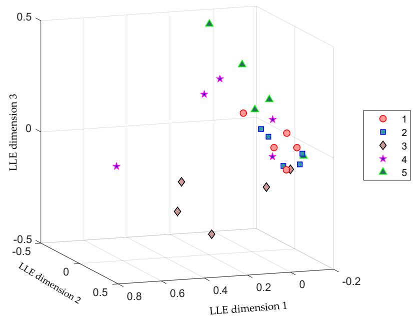

Figure 7. Three-dimensional scatter plot for the first three dimensions obtained using the LLE method on the juice data set.

4.2. Dimensionality reduction using LLE

The next step in the juice recognition methodology using a cyclic voltammetry electronic tongue is to reduce the dimensionality of the data. In this case, the Locally Linear Embedding (LLE) algorithm was used, which allows to carry out the feature extraction process. The results of the first 3 dimensions after applying the LLE algorithm to the juice dataset are evidenced in the scatter diagram of Figure 7. There is an overlap of classes 1, 2 and 3, which is why the use of a machine learning classifier algorithm is necessary. In this case, the classifying algorithm was k-Nearest Neighbors described below in the next section.

4.3. Classification and cross validation

The LLE algorithm needs the definition of the destination dimension, to find this parameter, a study of the change of the destination dimension d vs the classification precision obtained by the algorithm k-NN with $k = 1$ was carried out and Euclidean distance was considered. The cross-validation process executed was leaving one out (LOOCV) due to the small number of samples in the juice data set. As it can be seen in figure 8 the precision behavior tends to increase as d is increased up to a maximum of $d = 9$ for a classification precision of 80%. After the dimension $d = 9$ accuracy tends to decrease, in this sense the optimum size selection was defined as $d = 9$. Therefore, the size of the characteristics matrix at the input of the k -NN classifier algorithm is equal to 25×9.

Figure 9 shows the results of the confusion matrix for the mentioned precision of 80%. In this case, class 2 was correctly classified, there was 1 error for classes 1,3 and 4; finally, the class that was classified worst was class 5 with two errors. Overall of the 25 total samples, 20 were classified well and 5 badly.

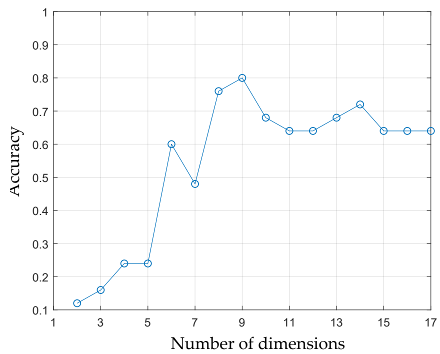

Figure 8. Behavior of the classification accuracy when varying the number of dimensions obtained with the LLE method at the input of the classifier algorithm k-NN.

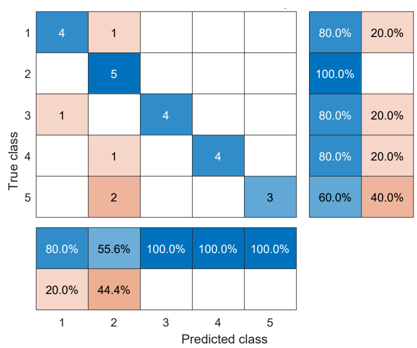

Figure 9. Confusion matrix for the juice data set obtained using 9 target dimensions of the LLE algorithm. The classification precision is equal to 80%.

5. Conclusions

This work presented a computational framework for processing the signals obtained by a cyclic voltammetry electronic tongue sensor array. The high classification accuracy obtained by the methodology in a dataset of 5 different juices showed the advantages of apply this methodology as classification method. It processes the raw complete voltammograms obtained by each working electrode and unfolded them to create a two dimensional matrix. This matrix was normalized applying the group scaling method. Then, the locally linear embedding method is used as a nonlinear feature extraction approach to obtain the feature matrix at the input of a k-N classsifier. As future work, the developed methodology will be applied for classify other kind of substances and other approaches related to semi-supervised classification. will be tested.

6. Acknowledgements

The authors thank FONDO DE CIENCIA TECNOLOGÍA E INNOVACION FCTeI DEL SISTEMA GENERAL DE REGALÍAS SGR. The authors express their gratitude to the Administrative Department of Science, Technology and Innovation—Colciencias with the grant 779—“Convocatoria para la Formación de Capital Humano de Alto Nivel para el Departamento de Boyacá 2017” for sponsoring the research presented herein. Jersson X. Leon-Medina is grateful with Colciencias and Gobernación de Boyacá for the PhD fellowship.

References

1. Podrażka, M.; Bączyńska, E.; Kundys, M.; Jeleń, P.S.; Witkowska Nery, E. Electronic tongue—a tool for all tastes? *Biosensors* **2018**, *8*, 3.
2. Ross, C.F. Considerations of the use of the electronic tongue in sensory science. *Current Opinion in Food Science* **2021**.
3. Zabadaj, M.; Ufnalska, I.; Chreptowicz, K.; Mierzejewska, J.; Wróblewski, W.; Ciosek-Skibińska, P. Performance of hybrid electronic tongue and HPLC coupled with chemometric analysis for the monitoring of yeast biotransformation. *Chemometrics and Intelligent Laboratory Systems* **2017**, *167*, 69–77.
4. Leon-Medina, J.X.; Vejar, M.A.; Tibaduiza, D.A. Signal Processing and Pattern Recognition in Electronic Tongues: A Review. In *Pattern Recognition Applications in Engineering*; IGI Global, 2020; pp. 84–108. doi:10.4018/978-1-7998-1839-7.ch004.
5. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *science* **2000**, *290*, 2323–2326.
6. Vitola, J.; Pozo, F.; Tibaduiza, D.A.; Anaya, M. A sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications. *Sensors* **2017**, *17*, 417.
7. Leon-Medina, J.X.; Cardenas-Flechas, L.J.; Tibaduiza, D.A. A data-driven methodology for the classification of different liquids in artificial taste recognition applications with a pulse voltammetric electronic tongue. *International Journal of Distributed Sensor Networks* **2019**, *15*.
8. Analog Devices (2021) EVAL-AD5940ELCZ Electrochemical Evaluation Board. https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/eval-ad5940elcz.html#eb-overview. Accessed: 2021-05-20.
9. Analog Devices (2021) EVAL-ADICUP3029. https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/eval-adicup3029.html#eb-overview. Accessed: 2021-05-20.
10. Leon-Medina, J.X.; Anaya, M.; Pozo, F.; Tibaduiza, D. Nonlinear Feature Extraction Through Manifold Learning in an Electronic Tongue Classification Task. *Sensors* **2020**, *20*, 4834.