Abstract

HyperTaste Lab—A Notebook with a Machine Learning Pipeline for Chemical Sensor Arrays †

Gianmarco Gabrieli * , **Michal Muszynski, Matteo Manica** , **Joris Cadow-Gossweiler and Patrick W. Ruch**

IBM Research Europe, 8803 Rüschlikon, Switzerland; mmu@zurich.ibm.com (M.M.); tte@zurich.ibm.com (M.M.); dow@zurich.ibm.com (J.C.-G.); ruc@zurich.ibm.com (P.W.R.)

* Correspondence: gga@zurich.ibm.com
† Presented at the XXXV EUROSENSORS Conference, Lecce, Italy, 10–13 September 2023.

Abstract: The cross-sensitivity of materials in low-selective sensor arrays, namely e-noses and e-tongues, results in a convoluted sensor array response, which renders traditional analytical methods for data processing ineffective. Machine learning approaches can help discover the latent information in such data, and various data processing methods, including unsupervised and supervised techniques, have been proposed to calibrate those devices. In this study, we demonstrate HyperTaste Lab—a notebook with a machine learning pipeline for potentiometric sensor arrays. The ability of the notebook to process raw data produced by model sensor arrays comprising cross-sensitive and/or ion-selective electrodes is demonstrated for the characterization of drinking water and consumer beverages. We describe the modular data processing and machine learning framework that can be applied by sensor researchers to accommodate different signal modalities and perform various downstream tasks, such as the verification of a product's originality, the estimation of ion concentrations, and the quantitative prediction of sensory descriptors.

Keywords: sensor arrays; electronic tongue; data processing; Jupyter Notebook

1. Introduction

Chemical sensors based on arrays of low-selective and highly sensitive sensors, such as electronic noses (e-noses) and electronic tongues (e-tongues), have shown potential to be used for fast and untargeted chemical analyses of multi-component media [1]. A key element that enables the practical use of those devices is often the analysis and interpretation of their combinatorial responses due to the inherent cross-sensitivity expressed by the array of sensing materials. Data analysis pipelines have been proposed to enable the processing of raw signals and transformations and the extraction of signal characteristic features [2]. Pattern recognition methods and machine learning approaches can then be leveraged to calibrate the sensor array and conduct a qualitative or quantitative analysis. However, setting up a complete pipeline for data processing can be extremely time-consuming and could prevent sensor experts from obtaining quick feedback on the quality of their own sensor hardware or on the possibility to leverage the sensor array for target use cases. In the present study, we provide a Jupyter Notebook [3] that allows for the automated processing of sensor array responses, including comprehensive data exploration as well as training, testing, and exports to appropriate formats of both classification and regression machine learning models.

2. Materials and Methods

An automated pipeline was built to process time series data recorded from an integrated array of potentiometric polymeric sensors [4,5] and packaged in a Jupyter Notebook (Supplementary Material). The main sections are shown in Figure 1 and include the following:

1. Loading data from a CSV file and splitting them into TXT files containing voltage data per sample;
2. The visualization of raw time series potentiometric data per sample;
3. Feature extraction [4,5] to reduce data dimensionality and batch effect correction;
4. A Principal Component Analysis (PCA) for data exploration and unsupervised analysis;
5. Supervised learning for classification and regression machine learning models;
6. The visualization of multi-output regression model predictions in radar charts;
7. The export of trained models along with model metadata in ONNX format.
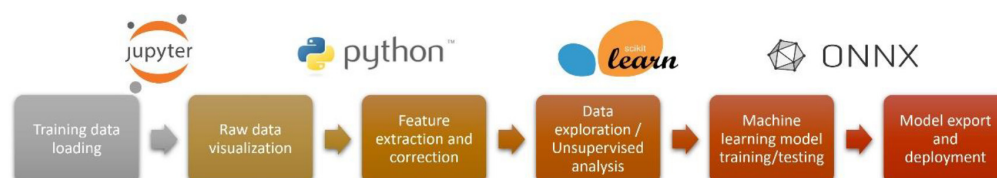
Figure 1. Machine learning pipeline for processing sensor array data in HyperTaste Lab.

The pipeline was applied to tests with drinking water for the estimation of Ca^{2+}, Mg^{2+}, and Na^+ concentrations as well as to coffee samples for the prediction of product originality and sensory profiles. Our framework makes use of common Python libraries, including *scikit-learn* and *onnxruntime*, for model training and export. The functionalities are integrated in the new Python package *hypertaste*.

3. Discussion

Establishing automated pipelines for sensor array data processing can accelerate sensor research. The present work provides an example framework that can be generalized to other transduction mechanisms, signal modalities, and target tasks. Furthermore, it makes machine learning techniques accessible to scientists for the initial exploration of novel approaches to enable the interpretation of complex sensor responses. Combining the ease of use of notebooks with common practices in sensor data processing results in a tool that speeds up end-to-end sensor development and testing.

References

1. Vlasov, Y.; Legin, A.; Rudnitskaya, A.; Di Natale, C.; D'Amico, A. Nonspecific sensor arrays ('electronic tongue') for chemical analysis of liquids: (IUPAC technical report). *Pure Appl. Chem.* **2005**, *77*, 1965–1983. [CrossRef]
2. Kirsanov, D.; Correa, D.S.; Gaal, G.; Riul, A.; Braunger, M.L.; Shimizu, F.M.; Oliveira, O.N.; Liang, T.; Wan, H.; Wang, P.; et al. Electronic Tongues for Inedible Media. *Sensors* **2019**, *19*, 5113. [CrossRef] [PubMed]
3. Perkel, J.M. Why Jupyter is data scientists' computational notebook of choice. *Nature* **2018**, *563*, 145–147. [CrossRef] [PubMed]
4. Gabrieli, G.; Hu, R.; Matsumoto, K.; Temiz, Y.; Bissig, S.; Cox, A.; Heller, R.; López, A.; Barroso, J.; Kaneda, K.; et al. Combining an integrated sensor array with machine learning for the simultaneous quantification of multiple cations in aqueous mixtures. *Anal. Chem.* **2021**, *93*, 16853–16861. [CrossRef] [PubMed]
5. Gabrieli, G.; Muszynski, M.; Thomas, E.; Labbe, D.; Ruch, P. Accelerated estimation of coffee sensory profiles using an AI-assisted electronic tongue. *Innov. Food Sci. Emerg. Technol.* **2022**, *82*, 103205. [CrossRef]