REAL-TIME ODOR CLASSIFICATION THROUGH SEQUENTIAL BAYESIAN FILTERING

Javier G. Monroy and Javier Gonzalez-Jimenez

Department of System Engineering and Automation, University of Málaga (Spain)

jgmonroy@uma.es, javiergonzalez@uma.es

ABSTRACT

The classification of volatiles substances with an e-nose is still a challenging problem, particularly when working under real-time, out-of-the-lab environmental conditions where the chaotic and highly dynamic characteristics of the gas transportation induce an almost permanent transient state in the e-nose readings. In this work, a sequential Bayesian filtering approach is proposed to efficiently integrate information from previous e-nose observations while updating the belief about the gas class on a real-time basis. We validate our proposal with two real olfaction datasets composed of dynamic time-series experiments (gas transitions are considered, but no mixture of gases), showing an improvement in the classification rate when compared to a stand-alone probabilistic classifier.

Index terms– Odor Classification, E-Nose, Sequential Bayesian Filtering.

1. INTRODUCTION

The detection of toxic or dangerous chemicals in human environments, the localization of multiple gas sources, or the generation of gas distribution maps in the presence of multiple chemical compounds, are just examples of applications that demand a reliable and fast classification of volatile substances. In all these cases, the classification system must be able to work under out-of-the-lab environmental conditions, which entails the absence of steady state signals in the gas sensor readings, as well as in real-time. It would be also highly desirable to provide some uncertainty measure about the class prediction (probabilistic classifier), since, for example, it may help in the decision-making process of a robot performing the odor classification.

Given the dynamic and chaotic nature of gas dispersal in real environments and the fact that e-nose data only contain information about the chemical substances at the time of the measurement, it is vital to integrate information from previous observations in order to obtain a reliable classification. Fig. 1 illustrates this fact with a real experiment of a naïve Bayes classifier which was trained to cope with four classes but was exposed to only one chemical substance. As can be appreciated, the posterior class probabilities fluctuate considerably along the experiment due to the dynamic nature of the e-nose signals, causing the maximum-a-posteriori (MAP) decision rule to incorrectly switch between different gas classes.

Up to date, different works have been proposed to solve this problem with relatively successful results [1] [2] [3]. In most cases, a reduced sequence of the complete olfactory time-series is used to extract a set of features that will be used later by the classification algorithm. Thus, these alternatives have to cope with the additional problems of data selection and feature extraction, which strongly influence the earliness and success rate of the classification method.

This work is supported by the Andalucía Regional Government and the European Union (FEDER) under research projects: TEP08-4016 and TEP2012-530.

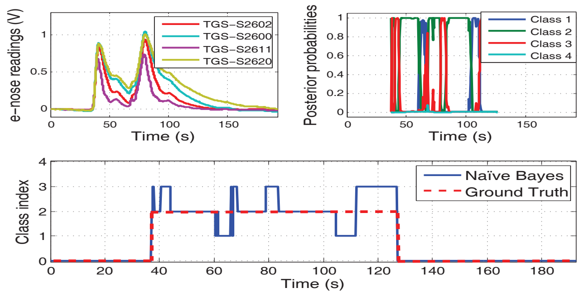

Figure 1. Instability of the classification due to the dynamic nature of the odor signal. Results obtained from a naïve Bayes classifier fed with the instantaneous response of an e-nose (after baseline manipulation) composed of four gas sensors.

In this work, we pursue a solution to the online classification of the dynamic and fluctuating data provided by an e-nose working in a real environment. The basis of our proposal is to smooth the outcome of a probabilistic classifier (exploiting the temporal correlation of the data) while keeping its earliness in delivering the optimal decision for each new measurement available. Concretely, we suggest applying sequential Bayesian filtering (SBF) to the posterior of any probabilistic classifier working with the instantaneous response of an e-nose. This approach enables an efficiently integration of information from previous e-nose observations without relying on data sequences, which in turns allows updating our belief about the chemical category in a real-time basis. As demonstrated in the experimental section, SBF also improves the classification rate respect the stand-alone probabilistic classifier working under similar circumstances.

2. SEQUENTIAL BAYESIAN FILTERING

The odor classification problem can be expressed as a hidden Markov model (HMM) where latent variables represent the class labels C, with $C \in (C^{(1)} \ldots C^{(M)})$, and the observed variables Z are the readings of the e-nose at each time step. Fig. 2 depicts the Bayesian network representing the conditional independence relations of such HMM. Naturally, this implies assuming the Markov properties, i.e. $P(C_t|C_{1:t-1})=P(C_t|C_{t-1})$ and $P(Z_t|C_t,Z_{1:t-1})=P(Z_t|C_t)$.

Then, our objective is to estimate the belief $Bel(C_t) = P(C_t|Z_{1:t})$ given we have two sources of information namely the posterior distribution $P(C_t|Z_t)$ of the selected classifier (which, as previously stated, can be any probabilistic classifier working on the instantaneous response of the e-nose), and the transition probability $P(C_t|C_{t-1})$ which specifies how the category states evolve over time. The latter can be defined as:

$$P(C_t|C_{t-1}) = \begin{cases} p_s & if\ C_t = C_{t-1} \\ \dfrac{1-p_s}{M} & otherwise, \end{cases} \quad (1)$$

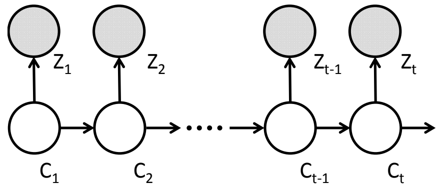

Figure 2.Bayesian network of the HMM for the real-time odor classification problem.

where M is the total number of class labels, and p_s the probability that two consecutive observations come from the same odor source, that is, that they are samples from the same class. In practice, this probability is very difficult to estimate, thus, it has to be set from the prior knowledge of the application. Consequently, we evaluate its influence on the classification performance in the experimental section.

We can therefore express our belief or overall posterior probability recursively over time, taking the form of a sequential Bayesian filter:

$$Bel(C_t) \propto P(Z_t|C_t) \sum_{i=1}^{M} P\left(C_t|C_{t-1}^{(i)}\right) Bel(C_{t-1}^{(i)}). \tag{2}$$

Note, however, that to sequentially estimate the $Bel(C_t)$, we require the conditional density $P(Z_t|C_t)$ whereas we have the posterior probability $P(C_t|Z_t)$. These two conditional distributions are related through Bayes' theorem (see [4]), which applied to (2) and omitting factors which are independent of the $\{C_n\}$ gives:

$$Bel(C_t) \propto \frac{P(C_t|Z_t)}{P(C_t)} \sum_{i=1}^{M} P\left(C_t|C_{t-1}^{(i)}\right) Bel(C_{t-1}^{(i)}), \tag{3}$$

where we consider that the marginal class probability $P(C_t)$ is time-independent, and thus, is learned from the training data. Consequently, the $Bel(C_t)$ at a given time step depends only on the posterior provided by the selected classifier, the class transition probability, and the belief at the previous time step.

3. EXPERIMENTS AND RESULTS

This section evaluates the performance of using SBF for the real-time classification of chemical volatiles in uncontrolled, real environments. As stated in Section 1, SBF can be applied to the output of any classifier which provides a posterior probability for each class. Nonetheless, for evaluation purposes, we restrict our selection to a naïve Bayes (NB) classifier because of its easy implementation and good performance as a stand-alone classifier. Two distinct datasets have been selected on the basis of real olfaction experiments with highly dynamic time-series: DS-UMA [3], and DS-UCI [5]. The former is based on an array of 6 MOX sensors for the detection of 4 different volatile substances, while for the latter, only a subset of time-series corresponding to the parameters L4, V_h=5, fan=100, down-sampled to 1Hz, and restricted to only 4 gas-classes is used.

Fig. 3(a) shows the classification performance obtained by 10-fold cross-validation for different values of the class transition probability parameter p_s. As can be appreciated, SBF improves the classification rate with respect to the stand-alone classifier (which corresponds to p_s=0.25 for a four-gas-class problem). A maximum improvement of 3.27% and 1.4% for DS-UMA and DS-UCI, respectively, is achieved for p_s=0.99, reflecting the fact that consecutive odor observations have a high likelihood to share the same chemical category.

An important limitation of both dataset when evaluating the performance of SBF is the absence of experiments where different volatiles are consecutively presented to the e-nose

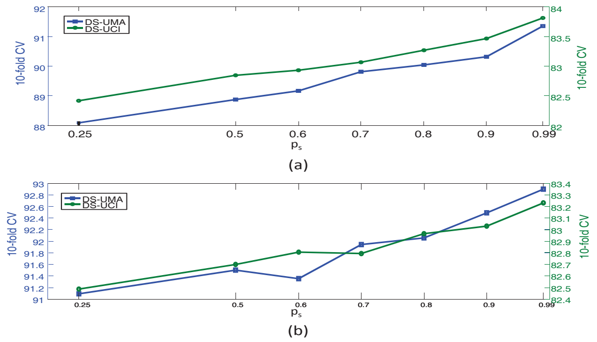

Figure 3. Classification rate obtained by a SBF working on the posterior of a naïve Bayes classifier for different values of the class transition probability. (a) Standard datasets, (b) datasets with class transitions.

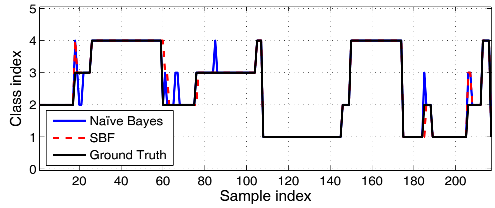

Figure 4.Example of classification with SBF of a time-series with class transitions.

(class transitions). This is due to the real complexity of obtaining a ground truth of such gas transitions when the gases are released in a dynamic real environment. Thus, we artificially simulate gas transitions by concatenating randomly selected sequences of time-series corresponding to different odor classes. Fig. 4 shows an example of the classifications results for such scenario, while Fig. 3(b) analyzes the SBF performance with respect to parameter p_s for this new configuration. Even when class transitions are considered, the benefits of SBF are noticeable, improving the classification robustness when considering dynamic and fluctuating e-nose data streams.

4. REFERENCES

[1] R. Menzel and J. Goschnick, «Gradient gas sensor microarrays for on-line process control - a new dynamic classification model for fast and reliable air quality assessment,»*Sensors and Actuators B: Chemical,* pp. 115-122, 2000.

[2] N. Hatami and C. Chira, «Classifiers with a reject option for early time-series classification,» in *IEEE Symposium on Computational Intelligence and Ensemble Learning,* 2013.

[3] F.-M. Schleif, B. Hammer, J. G. Monroy, J. Gonzalez-Jimenez, J. L. Blanco-Claraco, M. Biehl y N. Petkov, «Odor recognition in robotics applications by discriminative time series modeling,» *Pattern Analysis and Applications,* pp. 1-14, 2015.

[4] P. Komma, C. Weiss y A. Zell, «Adaptive bayesian filtering for vibration-based terrain classification,» in*IEEE International Conference on Robotics and Automation (ICRA),* 2009.

[5] A. Vergara, J. Fonollosa, J. Mahiques, M. Trincavelli, N. Rulkov y R. Huerta, «On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines,» *Sensors and Actuators B: Chemical,* vol. 185, nº 0, pp. 462-477, 2013.