

Chapter 1

Feature Selection and Sensor Array Optimization in Machine Olfaction

Alexander Vergara

BioCircuits Institute (BCI), University of California, USA

Eduard Llobet

University Rovira i Virgili, Spain

ABSTRACT

In the last few years, growing attention has been given to strategies for feature and sensor selection in multi-sensor systems for machine olfaction. The two main approaches consist of selecting the features extracted from the sensor response to be used to build a multivariate model, or selecting an optimal subset of factors; for example, principal components or latent variables. Selecting from the full set of features is challenging because there is considerable overlapping among them. Furthermore, features are affected by noise. However, methods based on selected features are interesting because the variables chosen carry direct and relevant chemical information; i.e., response time is connected to chemical kinetics. Therefore, these methods are expected to be robust toward the experimental conditions of each specific application. Unlike feature selection, factor selection uses the full set of variables, including noisy variables, to compute the factors before selecting from among them. The selection of an optimal subset of factors is not necessarily straightforward because the magnitude of an eigenvalue is not always a measure of its significance for the calibration.

DOI: 10.4018/978-1-61520-915-6.ch001

2 Feature Selection and Sensor Array Optimization in Machine Olfaction

Several variable selection methods have been reported as useful (Blum & Langely 1997, Guyon et al. 2006, Naes & Martens 1998, Sun 1995). These include deterministic methods such as forward or backward selection methods, correlated principal component regression analysis of weights resulting from multiple linear regression, branch and bound regression, and stochastic methods such as generalized simulated annealing or genetic algorithms. In most machine olfaction applications, it is usually out of the question to make an exhaustive search because it is a very time-consuming process, given the large number of variables to be considered for selection. Deterministic methods are, most of the times, greedy methods in which, once a choice has been made, e.g. the selection or elimination of a variable, this decision is never reconsidered. Such techniques can make a good selection with relatively few operations but can get easily trapped in a local optimum of the search space. Unlike deterministic methods, stochastic methods such as simulated annealing or genetic algorithms are more likely to find a global optimum in reasonable computational time. In the case of stochastic methods, the next point to be explored in a solution space is chosen by stochastic rather than deterministic rules, and no assumptions about the characteristics of the problem to be solved are needed. Therefore, they are normally more generally applicable. Although stochastic methods are useful for selecting features, it has been shown (Jouan-Rimbaud, Massart, & Noord 1996, Llobet et al. 2004) that the solution found should be investigated carefully because these algorithms do not prevent meaningless features, such as random non-relevant variables, from being selected.

In this context, the main objective of this chapter is to provide the reader with a thorough review of feature or sensor selection for machine olfaction. The organization of the chapter is as follows. First the 'curse of dimensionality' and the need for variable selection in gas sensor and direct mass spectrometry based artificial olfaction is discussed. A critical review of the different techniques employed for reducing dimensionality follows. Then, examples taken from the literature showing how these techniques have actually been employed in machine olfaction applications are reviewed and discussed. This is followed by a section devoted to sensor selection and array optimization. The chapter ends with some conclusions drawn from the results presented and a visionary look toward the future in terms of how the field may evolve.

INTRODUCTION

Machine olfaction applications would greatly benefit from gas sensors with high sensitivity and specificity or selectivity to target analytes, low cross-sensitivity to interfering species, fast response and full reversibility of the detection mechanism

59 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/feature-selection-sensor-array-optimization/52447

Related Content

A New Differential Evolution Based Metaheuristic for Discrete Optimization

Ricardo Sérgio Prado, Rodrigo César Pedrosa Silva, Frederico Gadelha Guimarães and Oriane Magela Neto (2010). *International Journal of Natural Computing Research* (pp. 15-32).

www.irma-international.org/article/new-differential-evolution-based-metaheuristic/45884

English to Hindi Machine Translation System in the Context of Homoeopathy Literature

Pramod P. Sukhadeve (2016). *International Journal of Artificial Life Research* (pp. 46-62).

www.irma-international.org/article/english-to-hindi-machine-translation-system-in-the-context-of-homoeopathy-literature/177184

Dual Hesitant Fuzzy Set and Intuitionistic Fuzzy Ideal Based Computational Method for MCGDM Problem

Akanksha Singh and Sanjay Kumar (2018). *International Journal of Natural Computing Research* (pp. 17-41).

www.irma-international.org/article/dual-hesitant-fuzzy-set-and-intuitionistic-fuzzy-ideal-based-computational-method-for-mcgdm-problem/214866

Immunocomputing for Spatio-Temporal Forecast

Alexander O. Tarakanov (2009). *Handbook of Research on Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies* (pp. 241-261).

www.irma-international.org/chapter/immunocomputing-spatio-temporal-forecast/19648

Linear Time Solution to Prime Factorization by Tissue P Systems with Cell Division

Xingyi Zhang, Yunyun Niu, Linqiang Pan and Mario J. Pérez-Jiménez (2011). *International Journal of Natural Computing Research* (pp. 49-60).

www.irma-international.org/article/linear-time-solution-prime-factorization/58066