

Variance-based selection of variables to improve prediction capability of innovative sensors in uncontrolled environment

Marine Dumon
COSYS/IMSE, Université Gustave Eiffel, Champs-Sur-Marne, France

Bérengère Lebental
COSYS/IMSE, Université Gustave Eiffel, Champs-Sur-Marne, France

Guillaume Perrin
COSYS/IMSE, Université Gustave Eiffel, Champs-Sur-Marne, France

Air and water pollution are now a major public health issue, and developing efficient detection methods to better monitor this pollution is essential to reduce the risk of exposure. Innovative materials such as nanomaterial-based sensors [1] have been proposed for their high sensitivity to different chemical species in air and water and their ability to detect them even in very low concentrations. However, if in laboratory these sensors are able to show encouraging results, the passage in real conditions generally pose difficulties. This is often due to the fact that these sensors are not very selective, and that in addition to reacting to changes in the concentrations of the pollutants of interest, they may also depend on several other environmental variables, such as temperature or relative humidity.

As the calibration relationship between the sensor outputs, the pollutant concentrations and the other environmental variables is often unknown, two distinct phases are needed for these sensors to be used in uncontrolled environments. In the first step, this relationship is estimated using labelled data provided by reference sensors. In the second step, this relationship is used to predict the pollutant concentration from the sensor outputs only. Several factors make this estimation challenging: the potential existence of unmeasured but influential pollutants, the measurement noise on the input and output data, and the likely non-linearity of the calibration relationship (see [2,3] for more details).

A key point of the calibration process is the selection of the appropriate environmental variables in the prediction model. In fact, due to strong correlations between environmental variables, one often observes that the calibration model performance improves when including a variable not directly influencing the sensor. On the other hand, this improvement actually constitutes “overlearning”, as it does not transfer to times or places where the correlations between environmental variables are different. Conversely, if one does not take into account an environmental variable that has a true influence on the sensor, one directly degrades its measurement performance. Identifying the optimal number of environmental variables to include in the calibration process for effective sensor deployment is therefore critical.

The present contribution focuses on this selection process. Based on experimental data, and simulated data (designed to be similar to the experimental one), we show that conventional sensitivity analysis techniques are confronted with considerable difficulties due to measurement noise and due to the high degree of correlation between variables. Similarly, we show that linear regressions and statistical tests allowing to identify the variables with very limited influence on the sensor responses are hindered by unobserved variables and do not fully answer the question.

As an alternative, we propose a new variance-based selection method. It allows a better compromise between significant influence on the sensor and noise levels for their measurement by auxiliary sensors.

References :

[1] Deepika Tyagi Huide Wang Weichun Huang Lanping Hu Yanfeng Tang Zhinan Guo Zhengbiao Ouyang Han Zhang. Recent advances in two-dimensional-material-based sensing technology toward health and environmental monitoring applications. *Nanoscale*, 12:3535–3559, 2020.

[2] Bérengère Lebental Guillaume Perin. Uncertainty-based calibration method for environmental sensors - application to chlorine and ph monitoring with carbon nanotube sensor array. *IEEE sensors journal*, 2023.

[3] Marine Dumon, Bérengère Lebental, Guillaume Perrin, Optimizing Sensor Calibration in Open Environments: A Bayesian Approach for Non-Specific Multisensory Systems, accepted to SMAI-JCM.

[Marine Dumon; COSYS/IMSE, Université Gustave Eiffel]
[marine.dumon@univ-eiffel.fr]