DOI 10.5162/SMSI2020/D5.1

An Approach for Feature Interpolation Between Classes to Simplify Calibration of Sensor Systems for Odour Monitoring

Julian Joppich, Manuel Bastuck[1], Andreas Schütze, Tilman Sauerwald[2]
Lab for Measurement Technology, Saarland University, Saarbrücken, Germany,
j.joppich@lmt.uni-saarland.de

Summary:

An approach for the augmentation of n-dimensional data was implemented and tested using artificial and real gas sensor data. With this method, a new (super)class is generated from two or more smaller classes by interpolating between these initial classes. It can be useful, for example, to assign a larger part of the feature space to a specific group of substances when performing classification tasks. Interpolation between two normally distributed classes generates very good results. For the interpolation between more than two classes or between inhomogeneous classes, several approaches were tested.

Keywords: class interpolation, features, classification, temperature cycled operation, MOS gas sensor

Background and Motivation

Model training is an important task for any sensor application. Prior to training, a sufficient data set has to be created, including reference data such as class assignments. Typically, a larger dataset can only be generated by conducting more measurements, which is time consuming and expensive, especially for gas sensors for odour classification. Moreover, for classification tasks of gas mixtures, the data set might still contain distinct gaps as only a limited number of gas concentration combinations can be tested efficiently. Especially if a group of gases is to be classified as one class and not identified individually, an approach to close these gaps and form a (super)class from two or more smaller classes might be desirable, which is the idea of the class interpolation presented here. We tested gases with similar odour, i.e. Sulphur compounds.

Interpolation Method

The interpolation is implemented as a MATLAB function based on a description of every class by its centroid and covariance matrix to represent the n-dimensional feature “cloud”. This implies that the classes are assumed to follow an n-dimensional normal distribution. Based on the centroids, the relation between two classes can be determined (distance and connection vector). The interpolated class is then built by successively generating normally distributed random clouds with a given centroid, covariance matrix and number of data points, with each of these variables being linearly interpolated between the two initial classes. An important parameter is the step length between the two classes. It is calculated from the extension of the individual class in the direction of interpolation weighed by the square root of the norm of the covariance matrices of the classes. For every sequentially generated cloud, the covariance matrix is reduced by a factor of two preventing the algorithm from generating too widespread clouds and extrapolating too far beyond the original margins. The size of the combined (super)class is reduced to max 10 times the size of the initial classes by transferring only every k-th artificial data point to the final array. This step reduces the computational expense if the algorithm is applied repeatedly and the risk of an unintended higher weight of the new class in relation to the initial classes.

Example Data

The experimental data used for testing the interpolation method were obtained from a gas measurement similar to [1]. A temperature cycled AS-MLV-P2 (ams Sensor Solutions Germany) was exposed to different mixtures of hydrogen sulphide (H_2S), dimethyl sulphide (DMS) and ammonia (NH_3) in varying hydrogen concentrations. The slopes of the logarithmic sensor conductance at low temperature plateaus were calculated according to the DSR model [2] with the toolbox DAV³E [3] and extracted as features. The cycle contains six low temperature plateaus, resulting in a six-dimensional feature space. As one gas class contains several concentrations, the classes are very inhomogeneous.

[1] Currently at WIPOTEC GmbH, Kaiserslautern, Germany.
[2] Currently at Fraunhofer Institute for Process Engineering and Packaging IVV, Freising, Germany.

Results

The interpolation between two normally distributed classes is always successful. However, if a class contains several sub-clusters, e.g. different concentrations of a gas, or if at least one of the input classes is of a more elongated shape, the resulting new class might have unexpected and not predictable shapes, sizes and densities. The same is true for the successive interpolation between three classes (interpolation between classes 1 and 2, followed by interpolation between the new class and class 3). In the case of three classes, it is more successful to first combine two classes to a temporary new class and then perform the interpolation with the third class. Fig. 1 depicts such an interpolation for three artificial random classes in 3D. The two classes at the bottom were combined first in one temporary class, interpolation of the temporary class with the class at the top resulted in the final class (cyan).

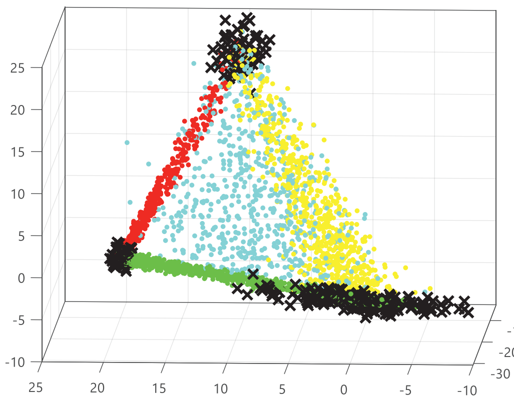

Fig. 1. Interpolation (cyan) between three artificial data clouds (black). The red, yellow and green data points represent the results of every possible two-class interpolation between the three initial classes.

The classes (features) of the two Sulphur compounds (H_2S and DMS) of the gas measurements were interpolated similarly. Instead of interpolating all H_2S and DMS concentrations directly, two temporary mixed classes were built by combining lower and higher concentrations of both gas classes, followed by an interpolation between the resulting two temporary classes, which lead to satisfactory results (Fig. 2).

Summary and Outlook

The implemented interpolation method leads to good results for all normally distributed classes tested and with some adjustments of the interpolation approach also for non-normally distributed classes. However, the robustness of the algorithm could be improved further and it needs to be tested on more datasets. Moreover, the performance of the interpolated (super)class for classification, e.g. when performing a linear discriminant analysis, will be tested.

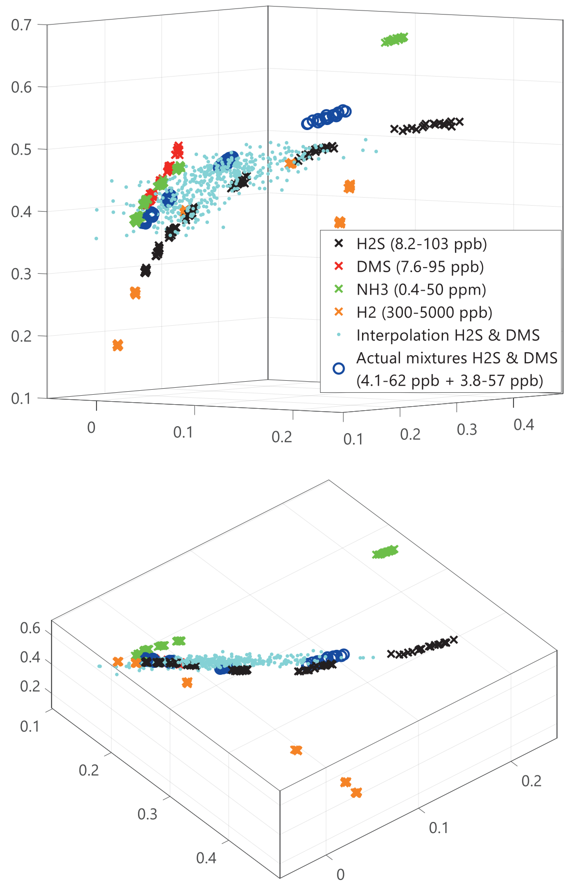

Fig. 2. 3-dimensional representation of the (6-dimensional) interpolation (cyan) between two classes (H_2S and DMS) from two perspectives. Individual gases and their interpolation form one plane, on which the actual mixtures of both gases (blue) lie also. Only a mixture with higher concentrations (highest blue data points) lies outside the interpolated class.

Acknowledgements

This work was carried out within the SEPEG project, funded by the Federal Ministry for Education and Research (BMBF) under project number 01IS17087. We thank our partners 3S – Sensors, Signal Processing, Systems GmbH, Saarbrücken, and Olfasense GmbH, Kiel, for the important contributions and the good cooperation.

References

[1] J. Joppich, A. Schütze, T. Sauerwald, Unterscheidung und Quantifizierung von Geruchsstoffen im ppb-Bereich mit temperaturzyklisch betriebenen MOS-Sensoren, *20. GMA/ITG-Fachtagung Sensoren und Messsysteme 2019*, 482-487 (2019); doi: 10.5162/sensoren2019/6.2.1

[2] T. Baur, C. Schultealbert, A. Schütze, T. Sauerwald, Novel method for the detection of short trace gas pulses with metal oxide semiconductor gas sensors, *J. Sens. Sens. Syst.* 7, 411-419 (2018); doi: 10.5194/jsss-7-411-2018

[3] M. Bastuck, T. Baur, A. Schütze, DAV3E – a MATLAB toolbox for multivariate sensor data evaluation, *J. Sens. Sens. Syst.* 7, 489-506 (2018); doi: 10.5194/jsss-7-489-2018