Contribution ID: **29**   Type: **not specified**

Binary Classification of Gas-Chromatograms using Data-Driven Methods

Monday, 27 June 2022 14:40 (20 minutes)

Gas chromatography (GC) plays an essential role in manufacturing daily operations for quality and process control, troubleshooting, research and development. The reliable operation of chromatography equipment ensures accurate quantitative results and effective decision-making. In many quality control and analytical labs, the operational procedure for GC analysis requires the chromatogram to be visually inspected by lab personnel to assess its conformity and detect undesired variation (e.g., baseline and peak shifts, unexpected peaks). This step is time-consuming and subjected to the experience of the observer, therefore, automating this task is crucial in improving reliability while reducing operational downtime.

Recent developments in data-driven modeling and machine learning have extended the relevance of these methods to a wide range of applications, including fault detection and classification. In this work, data-driven methods are applied to the task of chromatogram classification. Two classes of chromatograms are considered: a good class containing only expected variation, and a faulty class where upsets of different nature affect the quality of the chromatogram. Data-driven methods are built to distinguish between these classes, and both unsupervised methods (principal component analysis) and supervised methods (e.g., partial least squares discriminant analysis, random forests) are tested. The dataset utilized in this study was collected in a quality control laboratory and due to the low incidence rate of faulty GCs, chromatograms were simulated to increase the sample size and understand the impact of different fault types. The results indicate the successful detection of most types of faults and demonstrate the applicability of data-driven modeling for automating this classification task. Additionally, we highlight fault signatures (ghost peaks and broad peaks) that are more difficult to detect and require additional fine-tuning to be properly identified. The use of these models optimizes subject matter experts'time in handling chromatograms and improves the detection of unexpected variation in both the production process and the GC equipment.

Keywords

Smart GC, Machine Learning

Primary authors: RIZZO, Caterina (Eindhoven University of Technology/Dow);   RENDALL, Ricardo (Dow Inc.)

Presenter: RIZZO, Caterina (Eindhoven University of Technology/Dow)

Session Classification: CONTRIBUTED Modelling 2

Track Classification: Modelling