



Contribution ID: 46

Type: **not specified**

## **Imbalanced multi-class classification in process industries. Case study: Emission levels of SO<sub>2</sub> from an industrial boiler**

*Wednesday, 15 September 2021 14:00 (20 minutes)*

Imbalanced classes often occur in classification tasks including process industry applications. This scenario usually results in the overfitting of the majority classes. Imbalanced data techniques are then commonly used to overcome this issue. They can be grouped into sampling procedures, cost-sensitive strategies and ensemble learning. This work investigates some of them for the classification of SO<sub>2</sub> emissions from a kraft boiler belonging to a pulp mill in Brazil. There are six classes of emission levels, where the available number of samples of the highest one is considerably smaller since it reflects negative operating conditions. Four over-sampling procedures, namely SMOTE, ADASYN, Borderline-SMOTE and Safe-level-SMOTE, and the bagging (Bootstrap Aggregating) ensemble method, were investigated. All tests used an MLP neural network with a single hidden layer. The number of hidden units ([1:1:16]), the activation function (logistic, hyperbolic tangent), and the learning algorithm (Rprop, LM, BFGS), as well as the imbalance ratio, were also varied. The best results increased the AUC for the minority class from 83.9% to 93.6%, and from 80.4% to 89.1%, which represents a gain of about 10%, while keeping the AUCs of the remaining classes practically unchanged. This significantly increased the individual g-mean metric for the minority class from 60.9% to 79.8%, and from 52.9% to 76.3%, respectively, without significant changes in the overall g-mean metric, as desired. All results are given in average values. Imbalanced multi-class data generally appear in process industries, which claims the use of data imbalanced strategies to achieve high accuracy for all classes.

### **Keywords**

Process industry, Classification, Imbalanced data

### **Special/invited session**

**Primary authors:** Mr CARMO, Tomás (Federal University of Minas Gerais); ALMEIDA, Gustavo (Federal University of Minas Gerais)

**Presenter:** Mr CARMO, Tomás (Federal University of Minas Gerais)

**Session Classification:** Process 3

**Track Classification:** Process