



Contribution ID: 42

Type: **not specified**

## A semi-supervised approach to stream-based active learning for industrial processes

*Monday, June 27, 2022 11:00 AM (30 minutes)*

The high production rate of modern industrial and chemical processes and the high cost of inspections make it unfeasible to label each data point with its quality characteristics. This is fostering the use of active learning for the development of soft sensors and predictive models. Instead of performing random inspections to obtain quality information, labels are collected by evaluating the informativeness of the unlabeled instances. Several query strategy frameworks have been proposed in the literature but most of the focus was dedicated to the static pool-based scenario. In this work, we propose a new strategy for the stream-based scenario, where instances are sequentially offered to the learner, which must immediately decide whether to perform the quality check to obtain the label or discard the instance. The iterative aspect of the decision-making process is tackled by constructing control charts on the informativeness of the unlabeled data points and the large amount of unlabeled data is exploited in a semi-supervised manner. Using numerical simulations and real-world datasets, the proposed method is compared to a time-based sampling approach, which represents the baseline adopted in many industrial contexts. The results suggest that selecting the examples that are signaled by control charts allows for a faster reduction in the generalization error.

### Keywords

Active Learning, Statistical Process Control, Linear Regression.

**Primary authors:** CACCIARELLI, Davide (Technical University of Denmark (DTU)); KULAHCI, murat (DTU); TYSSDAL, John (NTNU)

**Presenter:** CACCIARELLI, Davide (Technical University of Denmark (DTU))

**Session Classification:** INVITED Young Statisticians inviting

**Track Classification:** Other/special session/invited session