Process Data Analytics: Theory to Practice

AIChE Spring Meeting, Houston, 2020

Bhushan Gopaluni, Fani Boukouvala, Ivan Castillo, Leo Chiang

Dec 11, 2019

Workshop Introduction

We are currently at the cusp of what is considered the fourth industrial revolution. This revolution is driven by the ubiquitous cyber-physical systems, algorithmic developments in artificial intelligence, gargantuan computing power, inexpensive memory and the gigantic volumes of data that are being collected. The process industries are in possession of treasure troves of heterogenous data that is gravely under utilized. The competitive global environment, and the ever increasing demands on energy, environment and quality are subjecting these industries to a high level of economic pressure. The incredible volumes of data that they already possess are poised to provide a level of automation and efficiency never seen before and thus alleviate the economic and competitive pressures.

Process industries have been using data analytics in various forms for more than three decades. In particular, statistical techniques such as principal component analysis (PCA), partial least squares (PLS) and canonical variate analysis (CVA) have been used widely. In addition, time series methods for modeling such as maximum likelihood estimation and prediction error methods have also been extensively applied on industrial data. The recent developments in machine learning and artificial intelligence provide an opening for using process data in new ways to extract information to solve modeling, process monitoring, fault detection and diagnosis and control problems. However, in order to successfully apply machine learning methods to process data, practitioners require not only a high level understanding of the algorithms but also strong programming in languages such as Matlab, Python and R as well as knowledge of packages such as TensorFlow, PyTorch, Keras and Jupyter.

This workshop will introduce the essential machine learning algorithms and software tools for graduate students, experienced researchers and engineers working in the industry. In particular, several known and emerging applications of these algorithms in soft sensing, state and parameter estimation, process monitoring, fault detection and diagnosis, and control will be presented. Specific focus will be given to formulation and solution of optimization problems with embedded Machine-Learning models (surrogate models), and adaptive model-based sampling. The workshop will include interactive hands-on exercises using Python and Google-Colab. Elementary knowledge of probability and statistics is required to attend this workshop.

Course Plan and Contents

This full day (approximately seven and half hours) workshop will be organized at the AIChE Spring Meeting. Starting with an elementary introduction to statistics and probability, we will develop various regression, classification, dimensionality reduction and advanced learning algorithms that are of interest to practicing engineers. In addition, various widely used machine learning software packages will be introduced. Registrants will receive take away software code to implement some of these algorithms.

1. Data-Driven Optimization
    - Data Pre-Processing
    - Regression
    - Design of Experiments
2. Classification & Regression Algorithms
    - k-nearest neighbours algorithm

- k-means algorithm
- Support Vector Machines & Logistic Regression
- Decision Trees
- Hierarchial Clustering
- Least Squares, kernels and regularization
- Applications in process and other related industries

3. Dimensionality Reduction Algorithms
    - Principal Component Analysis (PCA)
    - Partial Least Squares (PLS)
    - Discriminant Analysis
    - Nonlinear and dynamic dimensionality reduction algorithms
    - Applications in process and other related industries
4. Surrogate Modelling
5. Advanced Learning Algorithms
    - Artifical Neural Networks
    - Deep Learning
    - Reinforcement Learning
    - Applications in process and other related industries

Learning Outcomes

By the end of this workshop, registrants will be able to

- identify and solve classification, regression, dimensionality reduction and learning problems.
- work with softwares such as Python, TensorFlow, Keras.

Workshop Agenda

The sequence of workshop presentations will follow the tentative schedule below:

Start time	**End time**	**Presenter**	**Content**
9:00 AM	9:45 AM	Bhushan Gopaluni/Fani Boukouvala	Big Data Analytics - An Introducttion
9:45 AM	11:00 AM	Bhushan Gopaluni	Classification & Regression
11:00 AM	12:30 PM	Fani Boukouvala	Optimization Challenges with Process Data
12:30 PM	1:30 PM	N/A	Lunch
1:30 PM	2:15 PM	Bhushan Gopaluni	Dimensionality Reduction
2:15 PM	3:15 PM	Fani Boukouvala	Surrogate Modelling
3:15 PM	4:15 PM	Bhushan Gopaluni	Advanced Learning Algorithms
4:15 PM	6:00 PM	Ivan Castillo/Leo Chiang	Industrial Case Studies