CHE 4230 Advanced Process Control Systems

Ana B Terreros, Amanda M. Ross, and Teslim Olayiwola

Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, Louisiana 70803, United States

Introduction

The Eastman Chemical Co. uses a variety of process control technologies. Its Tennessee Eastman industrial chemical process (TEP) consists of five different unit operations with multivariable controls to produce liquid products G and H. Provided both the TEP manipulated and measured variables, the objective is to apply dimensionality reduction (DR), build a single layer artificial neural network (ANN) for the fault classification, and construct a support-vector machine (SVM) fault classifier. This report outlines our proposed approach and the timelines to achieve the project objectives.

Methodology

To uncover important patterns in the unlabeled TEP data, a quantitative research design will be employed using dimensionality reduction and clustering techniques. Herein, this part consists of two stages namely: the first stage will investigate the performance of at least five dimensionality reduction methods, such as principal component analysis, t-distributed stochastic neighbor embedding, and Linear Discriminant Analysis, along with at least three clustering methods, including DBSCAN and K-Means. The second stage will compare the ground truth label with the data mining results from stage one to assess the performance of the selected clustering algorithms.

Building on the results from part one, a single layer artificial neural network (ANN) and support-vector machine (SVM) will be used to create a fault classification model. Preprocessing procedures using different subsets of faulty and normal datasets will be considered. The performance of the ANN will be tested with different configurations of hidden-layer neurons and other necessary hyperparameters. Additionally, an SVM fault classifier will be built, and the performance of both models will be compared and discussed. To ensure robustness and generalization of the models, a 70:30 ratio will be used for training and testing, respectively. Furthermore, a fivefold cross-validation will be conducted on all models, and the model performance will be reported as the root mean squared error (RMSE). To execute the proposed methodology, multiple data analytics and machine learning libraries will be utilized. For machine learning, we will use *scikit-learn*[1] and *pytorch*[2]. In the data analytics phase, we will use *scikit-learn* and data handling libraries such as *pandas* and *numpy*. We will also incorporate additional publicly available libraries. Overall, this project aims to provide insights and establish a fault detection and classification model for the TEP data, which can have a significant impact on the industry.

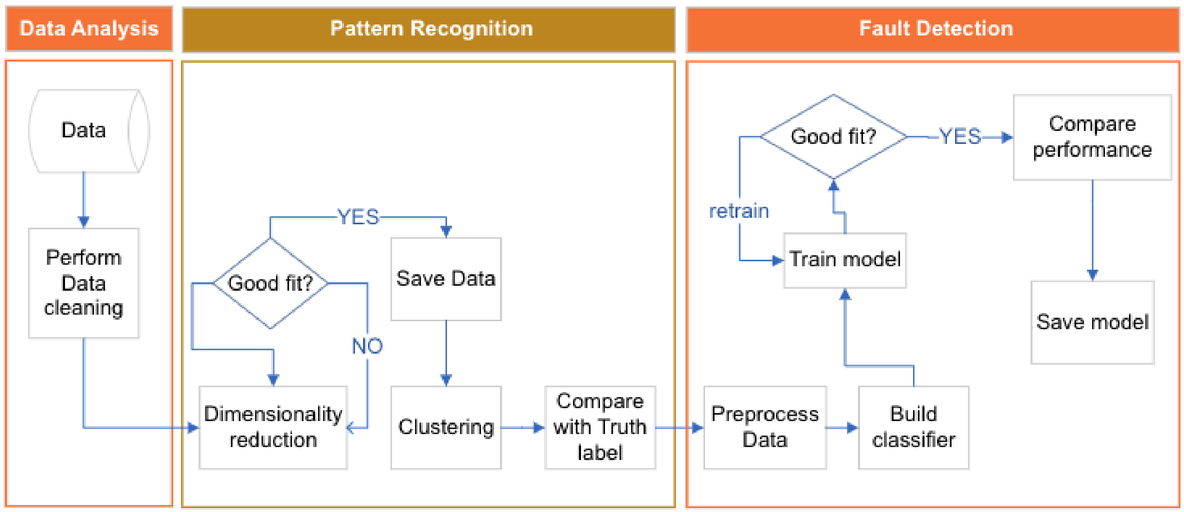

Figure 1. Workflow of the modeling strategies involved in this project.

Project Schedule

Task	Contributors	Tentative Deadline
PART 1: Create GitHub repository and prepare 2-page report	Amanda, Ana & Teslim	Feb. 17th
PART 2.1: Data preprocessing & Dimensionality reduction	Teslim, Ana	Feb. 26th
PART 2.1: Perform clustering and plot figures	Amanda, Teslim	Feb. 26th
PART 2.2: Create a 1-page summary of the main findings.	Amanda, Ana & Teslim	Mar. 10th
PART 3.2: Preprocess data datasets. Build ANN for the fault classification & assess the model.	Amanda, Ana & Teslim	Mar. 19th
PART 3.3: Build an SVM fault classifier.	Teslim	Mar. 21st
PART 3.4: Compare their performance and make a discussion.	Amanda & Ana	Mar. 23rd
FINISH PROJECT: review and submit	Amanda, Ana & Teslim	Mar. 24th

Note: the person/people in charge and dates could vary depending on the length of the tasks.

References

1. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
2. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. Preprint at https://doi.org/10.48550/arXiv.1912.01703 (2019).