AI Enhanced Soft Sensors

Group #11

Gabriel Mathias
Arman Hosseinsarraf
Chandrahasreddy Pola
Behnaz Sheikhi

Winter 2022 – University of Calgary

Introduction

Process control of a chemical plant is guided by the many sensors found throughout the plant. These sensors can measure physical properties such as pressure, temperature, flow rate, etc. These data are usually transferred into a database and logged. Based on these sensor readings, various actors such as operators, operation managers, plant managers, etc., make long- and short-term decisions. For example, an operator might want to open a valve if a pressure sensor reaches a certain value to avoid dangerous pressure build-up in a vessel, or a reliability manager might want to study the effect of various parameters on the operation of the plant etc. Sensors cannot directly measure certain properties. However, these properties can be estimated using simple or complex mathematical algorithms using a software simulator with the model of the plant. The term soft sensor is applied to a sensor that collects signals from other sensors as input to a software model and outputs a processed signal. For example, the presence of ice particles flowing through a gas pipeline is not something that a sensor can measure. But, by knowing the composition, temperature, pressure, and flow rate of the gas, it can be estimated. Thus, a soft sensor could detect the presence of ice particles flowing through a gas pipeline to guide operators or develop maintenance schedules.

Objectives and Motivation

The main objective of this project is to build an infrastructure that allows a user to connect real-world data to software models to create data that cannot be directly measured. Effectively this allows the user to build their soft-sensors into their database. These soft-sensor data can then control processes, guide decisions, and maximize profit. Also, soft-sensors could allow for less hardware or resources on-site and lower capital and operating costs through optimization of data gathering. Another important feature of a soft-sensor is that it can be predictive and use the exact mathematical correlations or AI, predict the occurrence of an event in the future, and take corrective actions before the event takes place.

General Scope of the Project

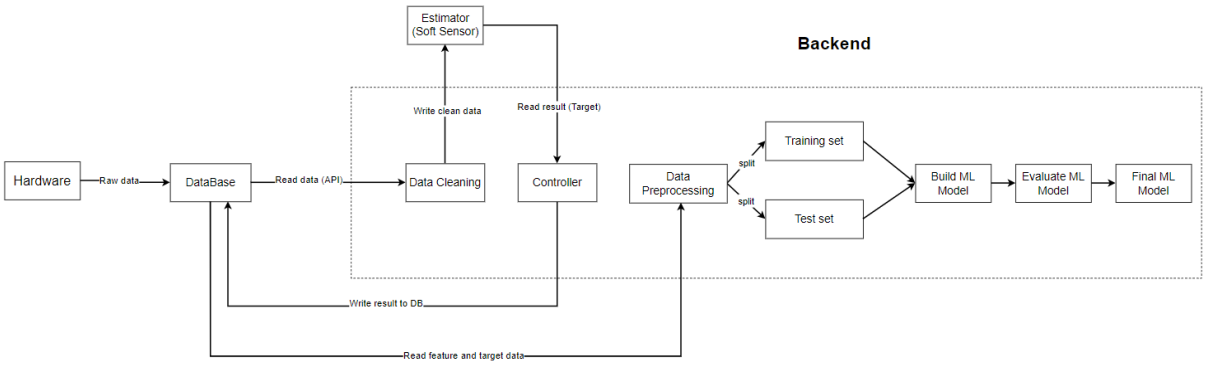
The general scope of the project is to develop an application that connects a source of plant data (database or real-time data source of sensor readings) to software models to provide estimates of properties that are not physically measured. The following are the main objectives of the application:

- Plant operating data will be sent to an appropriate application that calculates user-defined

properties from input data.

- Estimated properties must be pushed to a defined data destination such as a database.
- Data will be cleaned and preprocessed before sending it to the estimator application. A machine learning model will be trained with the input data and results from the estimator.

This project mainly focuses on the back-end development of the application and is based on the project supervisor's opinion, and the front-end development could be included in future steps. A conceptual diagram of the system is illustrated below:



Proposed Methodology and Approach

In this project, Python programming language and related existing libraries will develop the backend. In the first step, a pipeline diagram has been drawn to indicate how the entire system works and better understand the project's core. The data streaming part will be done through APIs and open-source software. In addition, the MQTT protocol will be studied to determine its suitability for direct sensor data transfer. Next, machine learning neural-based models will be trained after preprocessing the generated and collected data. The machine learning part of the project consists of four major steps:

- Cleaning and preprocessing the raw data
- Train the regression models
- Evaluate the models
- Optimize and choose the best model to be used in the backend.

If time permits, after finishing the steps mentioned earlier, which are the key and core of this project, we will design and build a User-interface for the application.

Main Challenges and Proposed Solutions

The list of challenges identified in accomplishing this project is as below:

- Communication between different applications and receiving and returning useful information on both sides would be one of this project's most important challenging parts. In order to do that, the APIs will be used for communication between the database and application.
- Selecting a Subscription protocol for data notification. The MQTT framework will be studied as a good candidate.
- Getting plant data for testing purposes. Use a process simulation software like UniSim to create plant data and add a noise function to simulate real-world plant data.

- Use the proper approach to clean and preprocess the data to extract useful information from the real world's original data, including noise. It can be done through the data analysis phase and the data preprocessing step using the python libraries such as pandas.
- Anomaly detection of sensors from the simulated data. This will be based on statistical data such as standard deviation.
- Tuning the hyper-parameters to get the best score when a machine learning model is applied. A grid search will be employed for the hyperparameters to address this issue.
- Designing the application with enough flexibility can be easily overloaded and used by different soft sensors, databases, etc.

Project Plan and Timeline

A GitHub repository has been created to let the team members contribute. Although the four team members are communicating through the slack channel and zoom meeting when needed, there will be weekly zoom meetings with the supervisor/advisor of the company. The start of the project will be on January 11th. Since the midterm presentation is on February 28th, the first draft will be completed in one week, and the slides will be finalized two days before the deadlines. The midterm's first version of the report will be completed one week before, and it will be finalized two days before March 7th. Since the final presentation is on April 11th, a first draft will be completed one week before, and the slides will be finalized two days before the deadline. Finally, the final report will be finalized on April 26th, two days before the deadline. The below list shows most of the process that will be done with their timeline briefly:

- Set up machines and GitHub
- Research available APIs to connect to databases (by February 28th)
- Develop infrastructure to connect to a database and read data (by February 28th)
- Develop infrastructure to write to the database (by February 28th)
- Generalize infrastructure to allow for any schema (by April 11th)
- Create communication between a software model and database (the user will give this software model) (by April 11th)
- Create a deep learning model that is trained on the background based on database inputs and software model outputs (by April 11th)
- If time allows, the following applications can be further developed:
    - Create a GUI
    - Implement other databases' technologies to be supported by the backend
    - Implement other estimator technologies to be supported by the backend