2021-2022
Second semester

University of Bahrain
College of Engineering
Department of Chemical Engineering

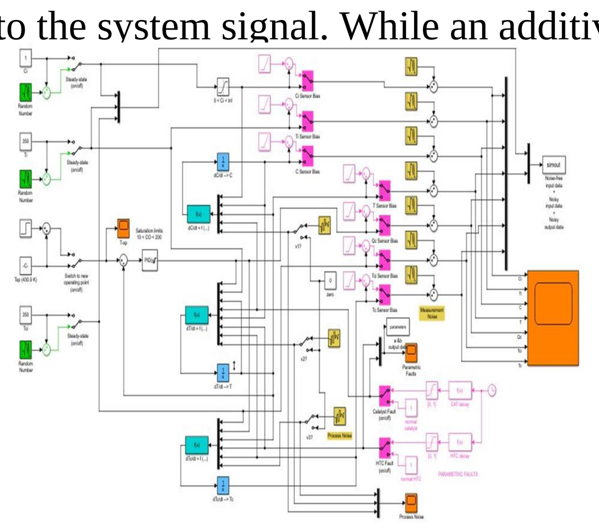
CHEMICAL PLANT BASED FAULT DIAGNOSIS USING RANDOM FOREST MACHINE LEARNING ALGORITHM

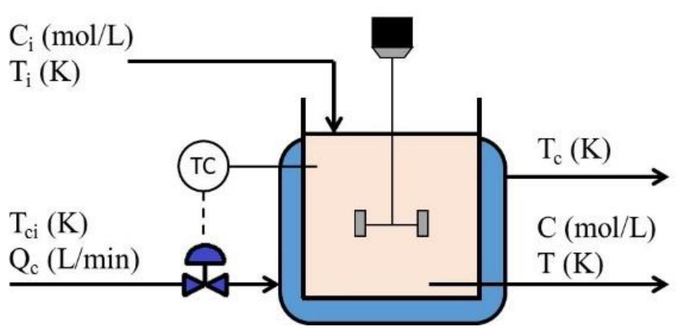
Abstract

This project is about diagnosing common faults that occur in chemical plants causing them to shut down or damage the equipment if it's not found early. An exothermic continuous stirred tank reactor was simulated using MATLAB to generate normal and faulty data sets that will be trained and tested by employing AI to create models using RStudio's coding tools. This technique enables fault location and detection in a matter of minutes, while the standard methods necessitate evaluating each piece of equipment and sensor, which is both time and money-consuming. The primary focus was on two approaches: RF and XGBoost, which were built in R programming languages, and afterward the data was introduced to both approaches to detect the faults. The data that was generated included ten faults induced inside the plant to be discovered and allocated by both approaches. The model performance has been evaluated by calculating the accuracy sensitivity, specificity, and PPR for two runs. XGBoost was the best fit, proving excellent detection performance for all sorts of faults from 90% to 100%, while RF produced less precise results, it is still considered relatively acceptable. Furthermore, for the future work since the field of artificial intelligence is wide and there are several techniques for detecting and diagnosing errors, it is better to discover new methods for detecting errors that enhance the total accuracy of the model to 100%, as well as to work on real industrial data to detect faults in it.

Design and Implementation

This section will introduce the R programming language and provide an overview of the reactor and reaction sorts employed in the plant which was used to gather the data. Finally, it involves recognizing and briefly describing each fault that has been brought to the plant, as well as expressing its influence on the output parameters over time. Additionally, the R programming language is a public, fully accessible, cross-platform software for data analysis, coding, and visualization of data that is supplemented by a huge userbase. Moreover, A closed-loop continuous stirred-tank reactor (CSTR) was created specifically for mimicking early-stage faults in the simulated plant. operating conditions are achieved by randomly disrupting the inputs around their nominal values every 1 hour. For all variables, the sample interval is 1 min. This plant is a nonlinear dynamic process with a total operating time of 20 hours (1200 minutes). The fault is established after 200 minutes of regular operation in faulty data sets. A single fault-free data set was performed to compare it to the faulty data, and the influence of each failure on the process parameters was investigated. Furthermore, A continuous stirred tank reactor (CSTR) is a batch reactor with a stirrer or other mixing device to achieve proper mixing. It is also a tank reactor with a continuous flow of reactants and reaction products. In an exothermic reaction, breaking bonds in the reactants requires less energy than the energy produced when new bonds form in the products. In exothermic reactions, the catalyst is utilized to lower the temperature and speed up the reaction without any losses in the process. As well, there are two types of faults: multiplicative and additive faults. A multiplicative fault multiplies an additional signal into the system signal. While an additive fault impacts the system signal by adding another signal to it.





Conclusion

This project emphasized the necessity of detecting persistent defects rather than unexpected breakdowns in chemical facilities. One of the most important ways for ensuring the safety of chemical processes is fault diagnosis. To recapitulate, although XGBoost and RF performed differently in detecting and diagnosing faults, the primary goal of this project was met. As the detection and diagnostic stage were readily acquired by utilizing AI machine learning rather than manually inspecting every piece of equipment and shutting down the entire facility. As a start, an extra step will be completed regarding the provided work. The data used in this research were gathered from a simulated plant, but it can gather genuine data from an actual running plant. Furthermore, supplemental enhancements would be implemented to both modeling techniques to enable them to operate online, which tends to mean that information will be gathered from the operating plant at the same moment, instead of offline, which indicates that data is to be collected after the plant had finally completed operating, achieving the greatest advantage of detecting and diagnosing a fault in the early stages.

Objective and Motivation

- **Motivation**

The fault diagnostic technique is essential since it might assist prevent occurrence spread and reduce the amount of equipment failure during abnormal events. Consequently, fault diagnosis helps in protecting the safety of employees and improves operational efficiency. It also plays a vital role in preventing environmental crises that might destroy the environment and violate pollution regulations, along with sparing the facility from a pricey full shutdown.

- **Aim of the project**

The purpose of this study was to identify ten various forms of faults that may occur in a simulator like such a plant with a continuous stirred tank reactor vulnerable to catalyst decay, as well as the potential of fouling phenomena and the type of reaming faults that could arise in sensors. If the tactic succeeds in collecting information to make it more realistic with the assistance of Simulink in MATLAB, it can be used in a practical implementation to entail faults in any plant by displaying the data to the model established with the aid of R programming language.

- **Constraints**

The primary challenges we experienced were a limited knowledge from the research stage through the finalization of the model and data collection, as well as time, because producing and gathering information takes time in addition to developing the model and validating its correctness. Due to time limits, we can only work in the offline stages, which means obtaining data from the plant after the operation has been finished rather than during it.

- **Standards**

This project focuses on the most crucial problem that chemical engineering plant operators face: fault diagnosis. As a result, one of the most significant tactics for guaranteeing the safety of chemical processes is fault diagnosis. Additionally, our project follows OSHA's alarm system rules, which assure worker safety and attentiveness in the case of an accident.

Results

XGBoost outperformed RF since it eliminated the multiplicative issue whilst simultaneously offering superb and class performance. By comparing their potential to recognize true positive values on run 2, the table below reveals that XGBoost is the preferred option. Table shows the percentage increase in TPR from XGBoost to RF. XGBoost has a higher chance that an actual positive prediction is positive (TPR), demonstrating the model's efficiency.

TPR%	RF	XGBoost	PERCENTAGE INCREASE OF XGBoost	PERCENTAGE ERROR
NORMAL	87.1%	98.9%	12.0%	13.6%
FAULT 1	48.7%	97.0%	49.8%	99.2%
FAULT 2	32.1%	100.0%	67.9%	211.5%
FAULT 3	59.2%	100.0%	40.8%	68.9%
FAULT 4	87.6%	100.0%	12.4%	14.2%
FAULT 5	95.5%	100.0%	4.5%	4.7%
FAULT 6	94.3%	100.0%	5.7%	6.0%
FAULT 7	78.2%	100.0%	21.8%	27.9%
FAULT 8	87.8%	100.0%	12.2%	13.9%
FAULT 9	90.6%	100.0%	9.4%	10.4%
FAULT 10	65.1%	98.8%	34.1%	51.8%
OVERALL	75.1%	99.5%	24.5%	32.5%

Student Name:
1- Samar Moawia Awadallah
2- Maram Rashid Rayash
3- Noor Ali Alkhadir

Student ID:
1- 20143850
2- 20164316
3- 20168246

Adviser Name:
1- Dr. Raed Al-jowder