# Soft Sensor Change Point Detection and Root Cause Analysis

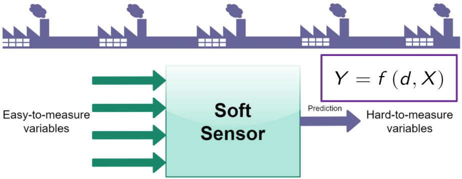
Liang Cao, Bhushan Gopaluni, Yankai Cao

The University of British Columbia

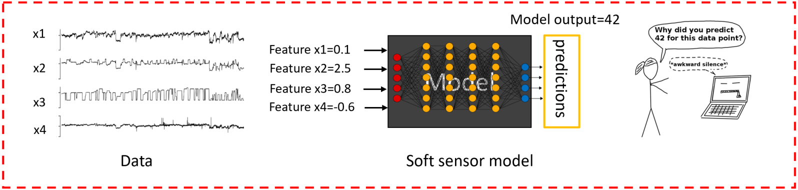
SICE 2022

## Soft Sensors

- A soft sensor is a mathematical model that uses measurable variables $X$ only to estimate the hard-to-measure state variables $Y$.
- For industrial soft sensors, getting predictions is not enough. It must also explain why it came to the model output and how to intervene when the model drifts.



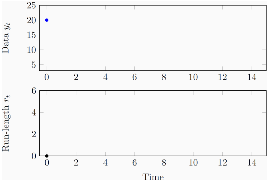
# Motivation

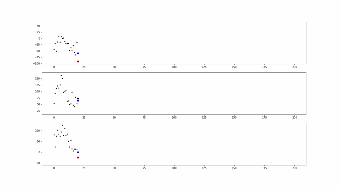



- How do we know if deployed soft sensor models are still performing well? Change Point Detection
- How to interpret the model result? Shapley value
- How to find the root cause of the change in soft sensor models? Shapley value+ Causal Discovery

# Change Point Detection

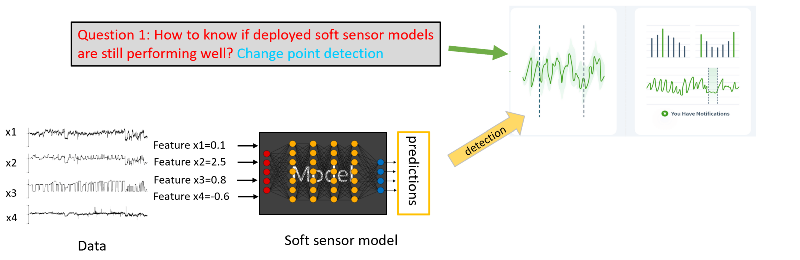
- Industrial processes are often non-stationary processes with constantly changing working conditions.
- Time-series segmentation; Partition into sub-sequences; Boundaries between partitions are change points (CPs).
- Change point detection (CPD) identifies structural/abrupt changes within a data sequence.



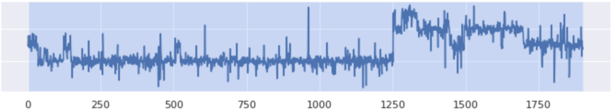
# Research Question 1

If there exists change points in soft sensor predictions, it indicates abrupt and significant changes in the model and may require us to rebuild the model to ensure that it does not drift.



# Change Point Detection

Input



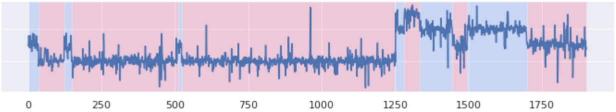

possible segmentations $\tau$

$$\min_{\tau} V(\tau, y) + pen(\tau) := \sum_{k=0}^{K} \sum_{t=t_k}^{t_{k+1}} \left\| y_t - \bar{y}_{t_k . t_{k+1}} \right\|_2^2 + \beta \left| \tau \right|$$

Change point detection

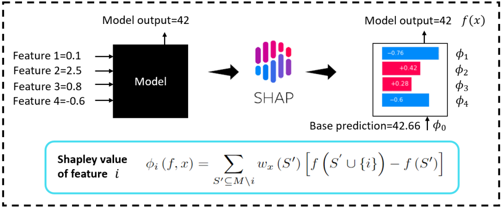
**Discrete optimization problem** ← **Search method** | **Cost function** | **Penalty**

Output



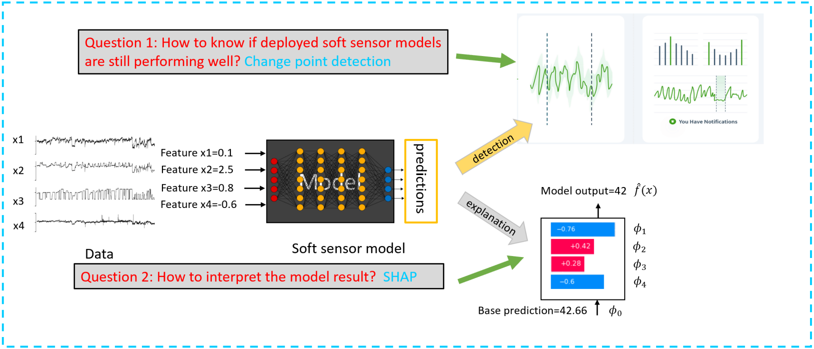
# Explainable model: SHAP (Shapley additive explanation)

- The SHAP value is the only machine learning explanation method that guarantees a fair distribution among the features.
- Explain the prediction of any machine learning model.
- Provide local & global explanation.



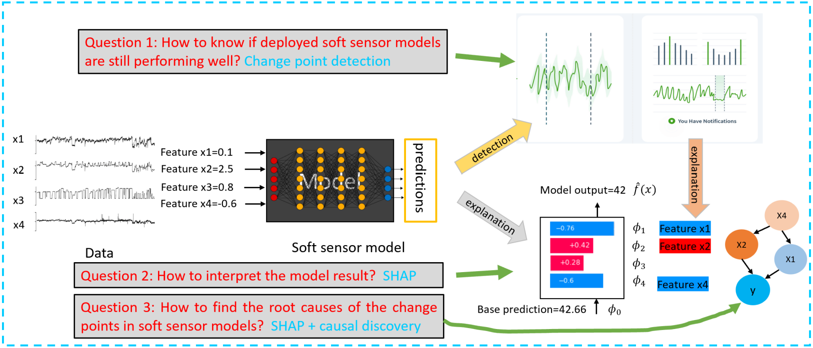
# Research Question 2

SHAP values can show how much each feature contributes, either positively or negatively, to the model predictions, including the entire data set (global) and each individual case (local).



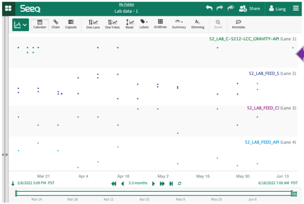
# Research Question 3

After a change point is detected, it is important to automatically locate the root cause of the change point. Causal discovery aims to infer causal structure from data.

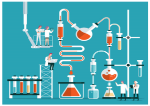


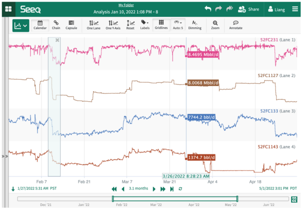
## Canada Parkland Refinery Data

- In this section, a real soft sensor data from Parkland refinery in Burnaby, British Columbia,Canada, is used for case study.
- 6 months of operational data from an FCC unit, 30 min sampling time, 10 process variables, 1 quality variable, 6586 samples.
- The objective is to accurately detect the change points and automatically find the root causes of changes.




Lab sample data

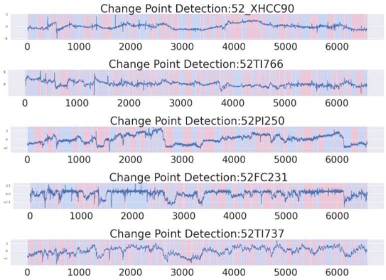


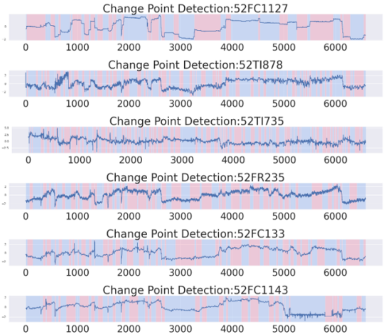



Process data

# Change Point Detection

Change point detection result for all features

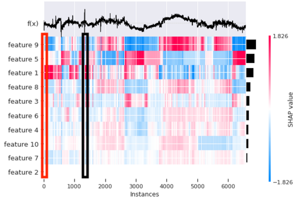


# Model Explanation

The local and global model explanation for the soft sensor modes

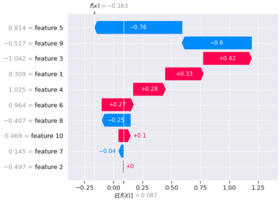

Global explanation:
feature importance



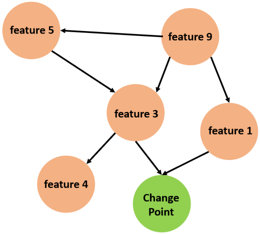

Local explanation:
feature attribution of all data

# Model Explanation

The local explanation and root analysis for change point 1260




Local explanation:
feature attribution of change point 1260




Causal discovery of change point 1260

- How do we know if deployed soft sensor models are still performing well?

✓ Change points of the soft sensors are detected by solving a discrete optimization problem.

- How to interpret the model result?

✓ Using SHAP to fairly distribute the contribution of each feature.

- How to find the root cause of the change in soft sensor models?

✓ Using causal discovery to find the root cause of change with change point data.

✓ The real-world commercial refinery case study validates the effectiveness of the proposed method.

# References


1. Ian Goodfellow, etc. Deep Learning, MIT Press, 2016.
2. Spirtes Peter, etc. Causal discovery and inference: concepts and recent methodological advances, Applied Informatics, 2016.
3. Renzhe Xu, etc. Why Stable Learning Works? A Theory of Covariate Shift Generalization, arXiv, 2021.
4. Charles Truong, etc. Selective review of offline change point detection methods, Signal Processing, 2020.
5. Liang Cao, etc. Data-driven dynamic inferential sensors based on causality analysis, Control Engineering Practice, 2020.
6. Liang Cao, etc. Causal Discovery based on Observational Data and Process Knowledge in Industrial Process, submitted.
7. Liang Cao, etc. Stable Learning with Small Samples, will submit.