Sensor Drift Detection and Estimation using Emerging Data Analytics Techniques

S. Cetiner, P. Ramuhalli, N. S. V. Rao, C. Greulich, P. Devineni (ORNL)
F. Zhang (UTK)
A. Gurgen (NCSU)
Motivation

• Leverage advances in data analytics and machine learning methods to address technical challenges in sensor drift detection, drift estimation, and uncertainty quantification

• The impetus is to build confidence in drift estimations and drift detections to enable sensor calibration extension
  – The sensor calibration extension is more of a regulatory impediment than a pure technical challenge
  – Development of analytical bounds may help strengthen the technical basis moving forward
Experimental Test Loop

Sensors considered in the fusion:

- Five differential pressure sensors:
  - Weed-D1Q
  - Foxboro-D3I
  - Rosemount-D4K
  - Safir-D2R
  - Weed-D5S
- One absolute pressure sensor:
  - Kulite-C1V
- Sensors were sampled at 20 Hz
Experimental Test Loop

Two transients are introduced in the loop:

- At $t=900 \text{ s (x=18,000)}$, loop temperature rises to $70^\circ \text{F}$
- At $t=3,600 \text{ s (x=72,000)}$, loop temperature rises to $90^\circ \text{F}$

The flow rate changes to maintain the loop temperature setpoints.
Scenarios

A multi-sensor dataset from an electrically heated flow loop provided by Analysis and Measurement Services (AMS) Corporation

Dataset includes twenty operational scenarios with various simulated failures/degradation conditions

- 2 normal operation (Scenarios 1 and 10; labeled AMS3 and AMS12)
- 8 simulated calibration changes (Scenarios 2-9; labeled AMS4-AMS11)
- 4 simulated blockages (Scenarios 11-14; labeled AMS13-AMS16)
- 3 simulated minor leaks (Scenarios 15-17; labeled AMS17-AMS19)
- 2 simulated air voids (Scenarios 18-19; labeled AMS20-AMS21)
- 1 simulated electromagnetic interference (Scenario 20; labeled AMS22)

We focus on drift estimation for Rosemout-d4k and Kulite-c1v
Diverse set of sensors with different ranges and calibration units

• Differential pressure sensors:
  - Weed-d1q: 0–850 in H₂O
  - Weed-d5s: 0–250 in. H₂O
  - Safir-d2r: 0–259 kPa
  - Foxboro-d3i: 0–50 psi
  - Rosemount-d4k: 0–750 in. H₂O

• Pressure sensor:
  - Kulite-c1v: 0–100 psi

Near identical measurements under no drift conditions

Three functionally replicated sensors:
  • Weed-d1q
  • Weed-d5s
  • Rosemount-d4k

scenario 1 (AMS3): training data
Training and test data sets

scenario 1 (AMS3): training data

scenario 2 (AMS4): test data

Rosemount-d4k – drifted by manipulating calibration
Two Machine Learning Methods for Multi-Sensor Fusion

- Support Vector Machine (SVM)
  - Gaussian Kernels
  - smooth estimator
  - non-linear mapping from input space into destination space

- Ensemble of Tree (EOT)
  - non-smooth estimator
  - mapping input dataset into collection of trees

Two regression estimators: very different designs
Multi-Sensor Fusion Method Using Machine Learning

- Differential pressure sensors:
  - Weed-d1q: 0–850 in. H$_2$O
  - Weed-d5s: 0–250 in. H$_2$O
  - Safir-d2r: 0–259 kPa
  - Foxboro-d3i: 0–50 psi
  - Rosemount-d4k: 0–750 in. H$_2$O

- Pressure sensor: Kulite-c1v: 0–100 psi

EOT and SVM methods:

- Regressions learned using training “non-drifted” measurements
- Tested using scenarios with externally induced drifts
Drift Estimates for Rosemount-d4k under Scenario 2

Ensemble of Trees (EOT)

Support Vector Machine (SVM)

error: expressed as percent of largest measurement
Drift Estimates for Rosemount-d4k and Kulite-c1v under Scenarios 11 & 14

**Rosemount-d4k: Scenario 11**

**Kulite-c1v: Scenario 14**
RMS Error for Drift Estimates for Rosemount-d4k under 20 Scenarios

- Root Mean Square (RMS) error is under 2.5% of largest Rosemount-d4k measurement
- EOT has lower overall error – under 2%
Comparison with AAKR

- Auto-Associative Kernel Regression (AAKR) is a nonparametric, memory-based modeling technique
  - measurements from a group of sensors are used to predict responses from all of them
  - residuals are used as drift estimates

Single sensor with large drift leads to large AAKR error
Drifted sensor measurements used for drift estimation
Performance Metrics

- **Time to detect drift onset**: the time at which the first true positive alarm is reported
  - For an ideal predictor, this should be 100%

- **Detection ratio**: the fraction of the detected drift/all the drift
  - For an ideal predictor, this should be 100%

- **False-alarm ratio**: the fraction of the false alarms
  - For an ideal predictor, this should be 0%

- **Scalability**: ability to support much higher number of sensors

- **Performance**: computational cost
  - Training cost
  - Inference cost

- **False alarm**: model erroneously indicates an error when there is none

- **Missed alarm**: a false negative where the model shows no indication of error despite an error being present
## Summary of advantages and disadvantages of investigated algorithms

<table>
<thead>
<tr>
<th>Method</th>
<th>Auto-associative?</th>
<th>Robustness</th>
<th>Training cost</th>
<th>Inference cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAKR</td>
<td>Yes</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>PCR</td>
<td>Yes</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>EOT</td>
<td>No</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>SVM</td>
<td>No</td>
<td>High</td>
<td>Medium to High</td>
<td>Low</td>
</tr>
</tbody>
</table>
Generalization error bounds were derived for the investigated techniques

• Generalization is one of the most important attribute of evaluating a learned model

• In statistical learning theory, generalization performance of a learning method relates to its prediction capability on a set of unseen samples drawn from the distribution same as that of the training set*

• Assessment of this performance is important in practice, since it guides the choice of learning method or model

Generalization error bounds were derived for the investigated techniques.

\[
\delta_{AAKR} = 8 \left( \frac{128 + 32h^2}{\varepsilon h^2 n_m e^{-\frac{2p}{h^2}}} \right)^{n_m p} e^{-\varepsilon^2 l/512}
\]

\[
\delta_{SVM} = 8 \left( \frac{32C}{\varepsilon} \right)^{2d} e^{-\varepsilon^2 l/512}
\]

\[
\delta_{EOT} = 16 \left( \frac{4l}{\varepsilon^2} \right)^{\left(1 + \frac{256BN_L}{\varepsilon} \right)} \log_2 \left( \frac{2el}{(\varepsilon+256BN_L)} \right) e^{-\varepsilon^2 l/2048}
\]

\(\varepsilon\)  Deviation bound between \(x_i\) and \(y_i\)

\(d\)  Input space dimension

\(C\)  Constant to satisfy Lipschitz property

\(l\)  Sample size

\(N_L\)  Number of leaves in tree
Summary

- Information/sensor-fusion method was used to estimate and detect sensor drift
- Two machine learning methods with different designs:
  - Smooth SVM method
  - Non-smooth EOT method
    - trained with “non-drifted” dataset
    - Testing using emulated testloop datasets: error with 2%
- The fusion methods were compared against the AAKR and PCR methods
- The methods generally exhibited performance improvements
Future Research

• Testing with plant data to show scalability

• Include drift detection problem; i.e., identify which sensor drifted?

• Improve generalization bounds on errors – a step towards uncertainty quantification of predicted sensor output
Acknowledgement

This work was supported under the Light Water Reactor Sustainability (LWRS) program.

We appreciate the support from AMS Co. for sharing the dataset.