Monitoring GOES-R ABI Radiometric Performance with a Machine Learning System

Zhenping Li, Ken Mitchell and Biruh Tesfaye
Arctic Slope Technical Services (ASTS)
Agenda

• Why Machine Learning solutions
• Machine Learning (ML) Approach to the systems with large number of detectors or sensors.
• Application in the GOES-R ABI radiometric performance monitoring
• Summary
Why Machine Learning (ML) in Radiometric Operations

• Instrument Radiometric Operation monitors the instrument calibration process,
  • Monitoring the radiometric performance in ground system
  • Assess the detector/sensor data quality.
  • Anomaly detection, and provide support in troubleshooting.

• Very Large number of datasets to be monitored
  • GOES-R ABI has over 7000 detectors comparing to total 16 detectors in GOES N-P Series.
    – The number of datasets to be monitored is over 20k.
  • The OPS Concept for GOES N-P Radiometric Operation no-longer works
    – Mostly manual process.

• Machine Learning approach automated the radiometric operations
  • Provides actionable information to engineers for paying close attention to specific datasets.
  • Enables very quick turnaround in troubleshooting in case of anomalies.
  • Provides the quality assessment of sensor/detector.
An instrument with large set of sensors/detectors is a dynamic system:

- Time dependent: the state variables $\{s_j(t_i)\}$ is a function of time
- Non-deterministic: datasets $\{d_j(t_i)\}$ are measurements of their state variable $\{s_j(t_i)\}$ with the Gaussian probability distributions.

ML creates the situational awareness of a dynamic system.
- A system with the ability to predict its own behaviors in the near future
- Dynamic monitoring compares the values of an incoming dataset with its predictions.

Post training analysis (clustering) separates the normal datasets from anomalous or low quality datasets based on the data training outputs.
- The data quality can be quantitatively characterized in ML framework.
A continuous dataset \( \{d_j(t_i)\} \) in a dynamic system is characterized by

- The time-dependent function:
  \[ s_j = f_j(t_i, \{s_k\}) \]
- And the noise level \( \sigma \):
  \[ \sigma_j = \sqrt{\frac{1}{N} \sum_i \left( f_j(t_i, \{s_k\}) - d_j(t_i) \right)^2} \]

The combination \( \{f_j(t_i, \{s_k\}), \sigma_j\} \) is defined as the time-dependent trend.

- The time-dependent trend is a representation of the Gaussian Probability Distributions.
- The time-dependent trend is an extension of the statistical trend used in the current operations.
  - If a state variable is independent of time, the time dependent function equals to its mean value.
Machine Learning algorithms generally include
- Data Model: \( F_j(t_i, \{S_k\}, W) \)
- Data training algorithm

Data Input: \( \{d_j(t_i)\} \)
- From telemetry, intermediate datasets in the instrument data processing, or any time dependent datasets from sensors or detectors.
- The datasets could be defective:
  - High noise level
  - Presence of outliers.

Outputs:
- Time dependent trends \( \{f_j(t_i, \{s_k\}), \sigma_j\} \).
- List of outliers \( \{O_j(t_i)\} \).

Data Training:

\[
\arg\min_W \sum_t \frac{1}{2} \left( D_j(t_i) - F_j(t_i, \{S_k\}, W) \right)^2
\]
- Finding the parameter set \( \{W\} \) to minimize the data model \( F_j(t_i, \{S_k\}, W) \)

Data training types:
- New training: the initial parameter set \( \{W\} \) is random numbers.
- Retraining: the initial parameter set \( \{W\} \) is known from datasets with the same data pattern.
• Time-dependent trending in operational environments is performed periodically in sessions to ensure that it captures both short-term data patterns and long-term changes.
• The trending period $T_0 = t_f - t_i$ must be long enough to capture the data patterns.
• Two neighboring trending sessions overlap to ensure the continuity and the stability of the trending outcomes.
• The output of the trending session $M$ is used as the input of the trending session $M+1$ to improve trending efficiency.
Dynamic Monitoring

Static Monitoring in current operations: Compare the value of a data point with the static red/yellow limits pre-defined in the telemetry database.

Dynamic Monitoring in the machine learning framework: Compare the value of a data point with the prediction of its time dependent trend. The data bound is determined by its noise level.
Anomaly Definition in Machine Learning

- An anomaly for a dataset in the machine learning is defined as the “unexpected data pattern changes”

- The data points of a dataset with Gaussian probability distribution satisfies

\[ |f_j(t_i) - d_j(t_i)| < N\sigma_j \]

- A data point outside of the data bound is defined as an outliers

- An outlier can be quantitatively characterized as

\[ O\left(d_j(t_i)\right) = N\left(\frac{f_j(t_i) - d_j(t_i)}{\sigma_j} > N\right) \]

- The data pattern change for a dataset corresponds multiple(persistent) consecutive outliers
  - Forms an outlier cluster.

- The outlier cluster is quantitatively defined as

\[ \chi_j^O = \sum_i \left(\frac{\delta_i^W}{T}\right) + \frac{N^E}{N^W} \sum_i \left(\frac{\delta_i^E}{T}\right) \]

- Used in real time monitoring and post training analysis.

- Monitoring and characterizing data pattern changes in the machine learning approach brings the fundamental advances in how the health and safety of a system with the large datasets are maintained.
Data Quality Metrics

• Data quality metrics measure data quality (or data pattern change) of datasets for a training session.
  • Quantitative and actionable.
  • Derived from the data training outputs

• Three metrics are defined:
  • Outlier clusters:
    \[
    \chi_j^O = \sum_i \left( \frac{\delta_i^W}{T} \right) + \frac{N^E}{N^W} \sum_i \left( \frac{\delta_i^E}{T} \right)
    \]
    \[0 \leq \chi_j^O < \infty\]
  • The temporal change measure the change in \(\sigma_j\) in the consecutive training sessions
    \[
    \chi_j^T = \frac{\sigma_j^M}{\sigma_j^{M-1}}
    \]
    \[0 \leq \chi_j^T < \infty\]
    • Significant increase in the metric may be caused by the data pattern changes.
  • The spatial change measures the relative data quality for a dataset group:
    \[
    \chi_j^S = \frac{\sigma_j}{\frac{1}{N} \sum_{k \in \{j\}} \sigma_k}
    \]
    • A dataset group is a set of datasets with the same patterns and scales
    • The mnemonics for detectors within the same spectral channels belongs to the same dataset group.
    • \[0 \leq \chi_j^S < \infty\].
Clustering analysis

- The distance between two datasets in the data quality space
  
  \[ d_{i,j} = \sqrt{(\chi_i^T - \chi_j^T)^2 + (\chi_i^S - \chi_j^S)^2 + (\chi_i^O - \chi_j^O)^2} \]

- The clustering technique:
  
  \[ \{S_j\} = \{S_j\}^N + \{S_j\}^E \]
  
  - \( \{S_j\}^N \): the main cluster with the majority of datasets that are normal.
  - \( \{S_j\}^E \): a collection of anomalous or poor quality datasets, generally do not form a cluster as each dataset has its own data quality characteristics.

- Define the center of main cluster \( \{S_j\}^N \) as \( \{\chi_N^T, \chi_N^S, \chi_N^O\} \), the distance to the center of main cluster as the aggregated data quality metric
  
  \[ d_{i,N} = \sqrt{(\chi_i^T - \chi_N^T)^2 + (\chi_i^S - \chi_N^S)^2 + (\chi_i^O - \chi_N^O)^2} \]

  - One define the threshold value for warning and error threshold values to determine the status of a dataset.
The ML Solution for ABI Radiometric Performance Monitoring

• Multi data models implemented:
  • Data Training algorithms should be very efficient in operational environment while maintaining sufficient accuracy.
  • Fourier Expansion Model, an extended Fourier Expansion Model, and Neural Networks
  • Use the linear model if it provides good description of datasets, as it is more efficient in data training

• The datasets monitored by ML system include the inputs and outputs of the calibration process,
  • The spacelook value for each detector,
  • ICT(black body) values for detectors in IR channel.
  • The gain parameters for each channel
  • Temperature data used in the IR channel calibrations.
  • Total number of datasets is over 20k.

• The post-processing to aggregate the metrics \( \{ \chi_j^T, \chi_j^S, \chi_j^O \} \) for each detector to show the operational status at the detector level.
  • Engineers only need to check the operational status for each channel; if there are issues, these metrics provide the detailed characteristics at the detector level
Variables in the Visible Channel

- Spacelook datasets use the neural network (4 nodes in the 1\textsuperscript{st} hidden layer, 2 nodes in the 2\textsuperscript{nd}).
- The data point spatial change $\chi_j^S$ value 2.4 corresponds to detector 360.
- The plot below represents the spacelook data points for detector 360 during the trending period. The metric provides a measure of the detector’s data quality.
  - The pattern is referred as the burst (or popcorn) noise.
  - May causes the occasional striping.
Variables in the Infrared Channels

- Gain data uses the neural network (6 nodes in the 1st hidden layer, 3 in the 2nd). Using the Fourier Expansion Model may lead to overfitting by including noise as part of the regular data pattern.
- Both spacelook_mean and ict_mean variables use the Extended Fourier Expansion Model to account for daily scale changes. This model has explicit linear dependence on time in its expansion amplitudes.
Both spacelook and ICT values became saturated around 364/10Z. The machine learning approach detected this immediately. The pattern change can be detected in real-time data monitoring, and also lead to very high values in both metrics \( \chi_j^T, \chi_j^S \). The saturated data leads to stripping in channel 12 images.

The pattern change cannot be detected with the existing approach using the static red/yellow limits.
An Example of Anomaly Detection with the metrics $\{\chi_j^T, \chi_j^S\}$

Both examples show elevated values for detectors 290 and 291.
The left column are hyper links (left graph) that enable users to navigate into detailed status of a specific channels (right).

- The detector ID column provides the hyperlinks for the plots of a specific detectors.
- Engineers only need to look at the status report for a given sessions.
- The data training and post training analysis are fully automated running as the scheduled tasks.
Summary

• The Machine Learning System brings the fundamental advances in how the health and safety of a system with a large number of sensors and detectors are monitored and maintained.
  • A dataset is monitored for its pattern changes, not static limit violations.
• Multi data models are used in operational environments for machine learning system to ensure the data training efficiency while maintaining the accuracy in training outcome
  • A machine system is a platform for machine learning algorithms
• The data quality metrics from data training output provide actionable information for engineers.
  • Use to quantitatively characterize the data pattern changes
• The rooting cause of an anomaly could be determined in minutes.
  • Significantly improves the system resilience.
• Can be used to systems with a large set of sensors/detectors.