Creating Situational Awareness with Spacecraft Data Trending and Monitoring

Zhenping Li
ASRC Technical Services, Zhenping.Li@asrcfederal.com
J.P. Douglas, and Ken Mitchell
ASRC Technical Services, JPaul.Douglas@noaa.gov and Kenneth.j.Mitchell@noaa.gov

Abstract

Data trending and monitoring are crucial in the spacecraft operations for maintaining the spacecraft health and safety and for evaluating the system performance and accuracy. The existing trend and data monitoring approaches are insufficient for time dependent datasets. The monitoring of these datasets would be very difficult if not possible without a time dependent trend being established, and the determination if a time dependent dataset at a given time is normal or in an error state requires significant engineering analysis efforts. This talk presents a Satellite Data Trending and Monitoring Toolkit(SDTMT), which implements a machine learning system for an automated and integrated trending and monitoring of time dependent datasets exhibiting the diurnal characteristics. Satellite data trending and monitoring are a natural fit to the operational concepts of a machine learning system. The data training in the machine learning system obtains the time dependent trend for datasets represented by the time function and standard deviation. The real-time or near real-time data monitoring determines if a data point is consistent with its time dependent trend. The potential anomalies can be detected in real time, which creates the enhanced situational awareness for autonomous spacecraft operations. The adaptive trending and limit monitoring algorithm and the neural networks are implemented in SDTMT as the machine learning algorithms. The machine learning approach is systematic, autonomous and adaptive. The application of the machine learning system to Geostationary Environment Operational Satellite(GOES) Imager data processing process is presented. It shows that the machine learning system enables the real-time monitoring of the instrument data calibration process that would have been impossible with the standard statistical trending approach. SDTMT can have many potential applications from the spacecraft health and safety to the science instrument data processing process, and it represents a significant advance toward an autonomous spacecraft operations.

1. Introduction

The spacecraft operation for maintaining the satellite health, safety, performance and accuracy involves data trending, data monitoring, and engineering analysis processes. The data trending is a data analysis process to determine a true measure of a dataset that is statistically distinguished from random behavior. The statistical approach has been a standard to the data trending, which the trend of a dataset is characterized by its statistical properties, such as the mean and standard deviation, for a given period. The data monitoring is performed in real-time by spacecraft telemetry and command system (TCS); the value of a data point is generally compared with a set of pre-defined and static limits to determine if it is normal, out of range, or in error state. The engineering analysis involves the targeted review of the specific dataset to identify, characterize, comprehend, and workaround an anomaly or a failure. The engineering analysis is generally a tedious manual process performed by engineers. The spacecraft operations have faced growing challenges from the new missions with more sophisticated onboard instruments and increased data volume. For example, the Imager on current NOAA GOES satellite has only 16 detectors, while the number of detectors for similar instrument for the new GOES-R mission has increased to thousands. The existing operational concept for maintaining the instrument to ensure its performance and accuracy is no-longer feasible for the new missions.

The focus of this paper is to develop and implement a machine learning approach for an automated and integrated approach to the data trending, monitoring, and engineering analysis for spacecraft operations. A machine learning system[1] consists of data models or algorithms that can learn from and make prediction on data. It generally has two stages; the data training stages to train a data model with a subset of data and the stage to predict the expect behavior for a dataset based on the outcome of data training stage. Instead of the traditional statistical data trending, the machine learning approach performs the time dependent trending, and most of spacecraft datasets are time dependent. The time dependent trend of a dataset also creates a dynamic limit to be used in the data monitoring, which has much tighter data bound than the traditional static limit. The dynamic limit of dataset enables the potential anomalies to be identified and characterized in real or near time, thus automates some of the engineering analysis process. This approach is predictive in the short term that enables the real time data monitoring and automated anomaly detection. It is also adaptive over the long term; the seasonal or long-term changes to datasets are captured automatically through the data retraining, and the threshold of data bound is determined by the noise level in datasets. As the system performance degrades, the noise level increases so that the data bound increases. This approach could have potentially wide range applications, which include the trending and monitoring of the spacecraft telemetry for health and safety and instrument data processing status.

The situational awareness is the ability of a system to perceive, comprehend, and make prediction of the data in its environment, thus, potential anomalies that deviates from the predicted behaviors could be automatically detected. The situational awareness of a system is crucial for a autonomic ground system[2], as it provides actionable data or information for an intelligent decision system or engineers to respond any potential problem accordingly. The machine learning presented here for data trending and monitoring provides a systematic approach for creating situational awareness. Once a machine learning system is setup and goes through the initial training to establish the data pattern in its data model, no further intervention is needed so that the data trending and monitoring become autonomous. The system will retrain itself with new datasets periodically to capture the long term or seasonal changes, and the data monitoring is using the most recent data training results. Anomalies can be detected automatically in real-time for engineers or an intelligent decision system for appropriate actions. A system with the situation awareness allows engineer to concentrate on the datasets with potential problems to enable a much more proactive operations. Early detections of potential anomalies minimize their impact to spacecraft operations.

The multi-layer feed-forward and back-propagate neural networks[3] are implemented as the machine learning algorithm. The data training algorithm implemented with the neural networks is shown to be systematic, accurate, adaptive, and efficient. The machine learning system presented here is similar to the Adaptive Trending and Limit Monitoring Algorithm (ATLMA)[4], in which the time function is expressed as a Fourier expansion. The least square fitting procedure is used in data training to obtain the coefficients in a Fourier expansion. The Fourier expansions obtained from the least square fitting are used in data monitoring. Using the Fourier expansion for a time dependent dataset has its limitations; Fourier expansions provide good descriptions to a limit set of data patterns, in which the long wavelength components in a Fourier expansion dominate. The neural network implemented in the machine learning system is much more adaptive to different data patterns. It has been shown that the neural networks with two hidden layers can be used to describe an arbitrary functions[5].

This paper is organized as the following. Section 2 provides a mathematical definition for data trending and monitoring of time dependent datasets. Section 3 discusses the data training approach in the machine learning system, which includes two training stages that follow very different training strategies. The neural network implementation is also discussed in Section 3. Section 4 shows how the machine learning system is applied to GOES imager calibration process. Section 5 provides a brief discussions on the software implementation of the machine learning system and general operational concepts when the software is integrated into a ground system. Finally, the summary is provided in Section 6.

2. Trending and Monitoring Time Dependent Datasets

The trending for time dependent datasets extends the standard statistical trending approach. A time dependent trend for a dataset$\{d(t_i)\}$ is represented by a time dependent function $f(t)$ for its true measure at a given time *t* and the standard deviation σ_e for its random behavior:

$$\sigma_e = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big(d(t_i) - f(t_i)\big)^2}. \qquad [1]$$

The function *f*(*t*) is finite, however, it may not be continuous. The value of a data point during the trending period should be within the range

$$|f(t_i) - d(t_i)| < N\sigma_e \qquad [2]$$

where the factor *N* is an integer defined by users. Once the factor *N* is set, it does not change for the lifetime of a system that generates the dataset. The valid data range in Eq. 2 is determined by the noise level, which could become larger if the system performance degrades. If a data point with the value outside the range defined in Eq. 2, it is defined as an outlier indicating a potential anomaly for the corresponding system.

The data training in a machine learning system is to find a time dependent function *f*(t) so that the error function

$$e = \frac{1}{2}\sum_i\big(d(t_i) - f(t_i)\big)^2 \qquad [3]$$

is minimum. For the special case of a constant function *f*(t), it can be shown that the function *f*(t) becomes the mean value of a dataset $\{d(t_i)\}$, and σ_e becomes the standard deviation in the statistical calculation. Therefore, the standard statistical approach can be regarded as a special case for the trending of time dependent datasets.

3. Data Training Approach in the Machine Learning System

An effective data training approach for data trending should meet the following requirements:

- It should be systematic for datasets with arbitrary scales and data patterns.
- It should be accurate to capture the complexity of data patterns so that the data-training outcome is close to its true time dependent trend for a dataset.
- It should be adaptive to the long-term changes of a dataset.
- It should also be efficient to minimize the impact on the computing resources.

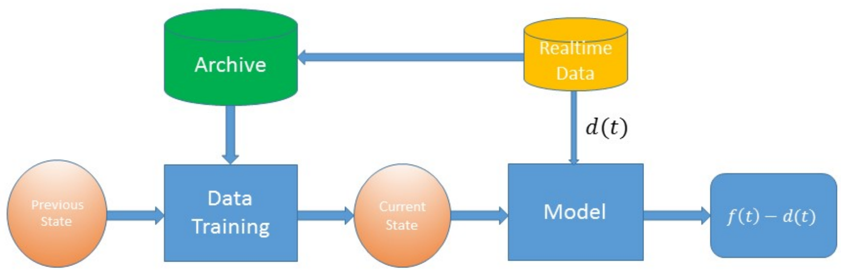

Figure 1 Data Training Approach

Figure 1 shows the data flow of the data training in a machine learning system for the spacecraft data trending and monitoring. The real-time data represent the telemetry data from the spacecraft that have been de-commutated by the ground system, the short term trend generated by TCS or a trending tool, and the intermediate data generated in a science instrument data processing system. The data archive in Figure 1 is generally part of a spacecraft ground system and ingest the real time data for either short or long term storage. The current and previous states in Figure 1 are represented by state variables, and a set of state variables is defined as a complete and sufficient representation of machine learning algorithm to characterize the time dependent trend $\{f(t), \sigma_e\}$. State variables can be archived and retrieved to reconstruct the time dependent trend, which are essential for the software implementation of a machine learning system. During the normal operations, the data training inputs the dataset from a data archive and state variables from the previous training period to generate new state variables. The data monitoring constructs the time dependent trend $\{f(t), \sigma_e\}$from the most recent state variables, and compares the incoming real-time data with the time dependent trend $\{f(t), \sigma_e\}$ to determine if the value of a data point is within the bound defined by Eq. 2.

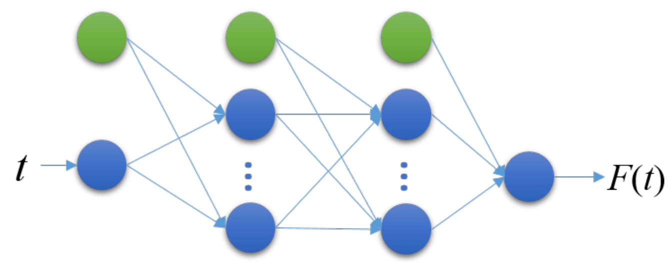

Figure 2 Two Hidden Layer Forward Feed Neural Network

The data training for the machine learning system has two phases: the initial training and retraining phases corresponding to the software deployment and operational phases. When a machine learning system is deployed into an operational environment, the system has not prior knowledge of patterns of datasets. The data training in this phase is to establish the structure of the machine-learning algorithm to capture the complexity of expected patterns of datasets based on the system performance requirements. Once the structure of a machine-learning algorithm is established during the initial training, it remains unchanged. The data retraining happens during the operational phase to capture the seasonal or long-term changes to data patterns. The machine learning system in this phase has the prior knowledge of datasets. The previous state variables shown in Figure 1 are used as the starting point for the retraining, and the difference between the retraining output and the previous state variables are expected to be small. Thus, the computing resource needed for data retraining is much less than that in the initial data-training phase, and this makes the data training more efficient during the normal operations.

The multi-layer feed-forward and back propagate neural networks are implemented for the machine learning algorithm, which has been proven to be much more adaptive to different data patterns. The network structure for the data trending and monitoring is shown in Figure 2. There are two hidden layers in the network structure. Both input and output layers contain only single node, which corresponds to the input time *t* and the output function *F*(t). The structure of a network with two hidden layers in Figure 2 can be represented by an integer array $\{1, n_1, n_2, 1\}$ with n_1 and n_2 representing the number of neurons at the hidden layer 1 and 2 respectively. The blue nodes in Figure 2 corresponds the network nodes, and the green nodes represent the bias nodes in a neural network. Thus, the data training for the network structure in Figure 2 becomes a curve fitting problem, and there have been extensive studies in the literature using neural network to do the curve fittings[6] in different applications.

The data training for the multi-layered neural network is to minimize the error function defined in Eq. 3, which is generally a nonlinear least squared fitting problem. During the initial data-training phase, the Levenburg-Marquardt (LM) back-propagation algorithm[7] is implemented in the data training at the initial phase, which has been proven to be accurate in capturing the complexity of the data patterns. However, the LM algorithm is dependent on the initial conditions, and it does not always converge especially for the initial random number due to the fact that it needs to evaluate the inverse of Jacobian matrix. Thus, the initial training implements a two-step approach, the simple gradient decent is used as the first step to obtain a set of weight parameters that are close to the solutions, and the resulting weight parameters from the first step are used as the input for LM back-propagation algorithm in the second step. Our simulation shows that this approach enables the data training to find the best solution very quickly. In the retraining stage, the data patterns are known for a given data set, and the daily changes to state variables are small. The gradient decent or its variation, such as the adaptive gradient decent[8], is more appropriate in the data retraining stage, which is more efficient.

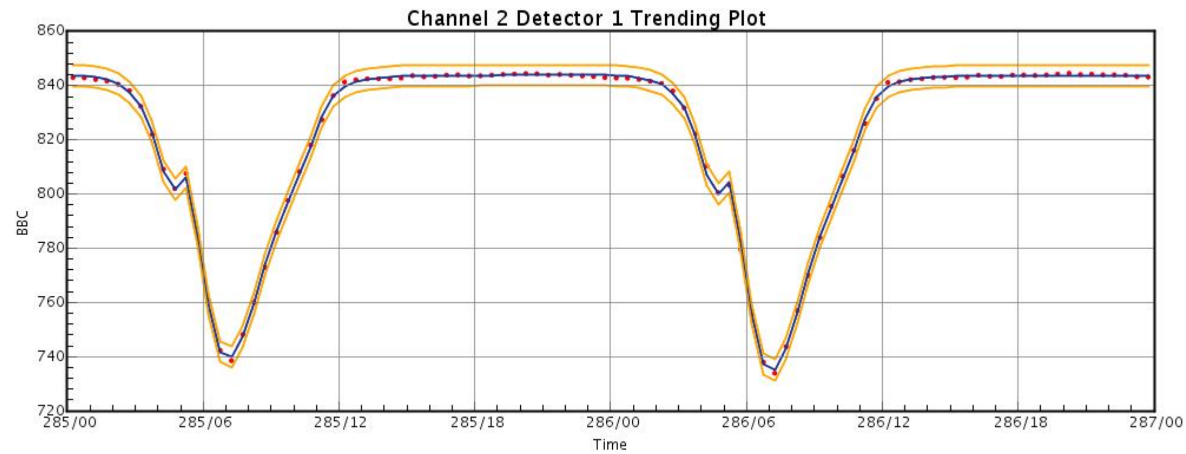

Figure 3 The output of the neural network data training for the raw black body datasets for detector 1 in channel 2. The blue line is neural network outputs, and the rad dot is the actual data from a short-term data archive. The two orange lines are the upper and lower limit value defined in Eq. 2.

4. Application to the GOES Imager Calibration Process

To show how the machine learning system for data trending and monitoring works, we present the results for the trending and data monitoring of the radiometric variables in GOES Imager data processing process. GOES Imager is a five (one visible and four infrared) channel radiometer designed to sense the radiant and solar reflected energy from sampled areas of Earth. There are 8 detectors for the visible channel and up to 2 detectors for each infrared channel. The Imager data processing in GOES ground system includes the data calibration and image navigation and registration to generate geo-located radiance images from the raw instrument data. The radiometric variables are the intermediate products from the instrument calibration process, which provides crucial insights into the instrument performance on the radiometric accuracy. The main inputs from the instrument calibration process in infrared channels are the spacelook data used as the calibration baseline, the blackbody data and instrument temperatures that provide a reference correspondence between the raw counts in an infrared channel and the temperature. The main outputs of the instrument calibration process are the bias and gain parameters, which are used to convert raw image pixels to its corresponding radiance. Both inputs and outputs from the instrument calibration processes are trended and monitored by engineers for possible signs of problems.

Figure 3 shows the data training results of the blackbody counts for the detector 1 in the infrared channel 2. The number the network nodes in the first and second hidden layers in Figure 2 are 4 and 2 respectively. The blue line represents the output of neural networks from the data training in the initial training phase. The two orange lines represents the upper and lower bound defined in Eq. 2, and the parameter N is 5 in this case. The training datasets covers 2 days, which show a very good diurnal behavior. Figure 3 shows the time dependent trend $\{f(t), \sigma_e\}$ generates a much tighter bound defined by its standard deviations, which would have been impossible for static upper and lower limits.

Therefore, Figure 3 provides a visual approach to the data monitoring based on the time dependent trend from the machine learning system.

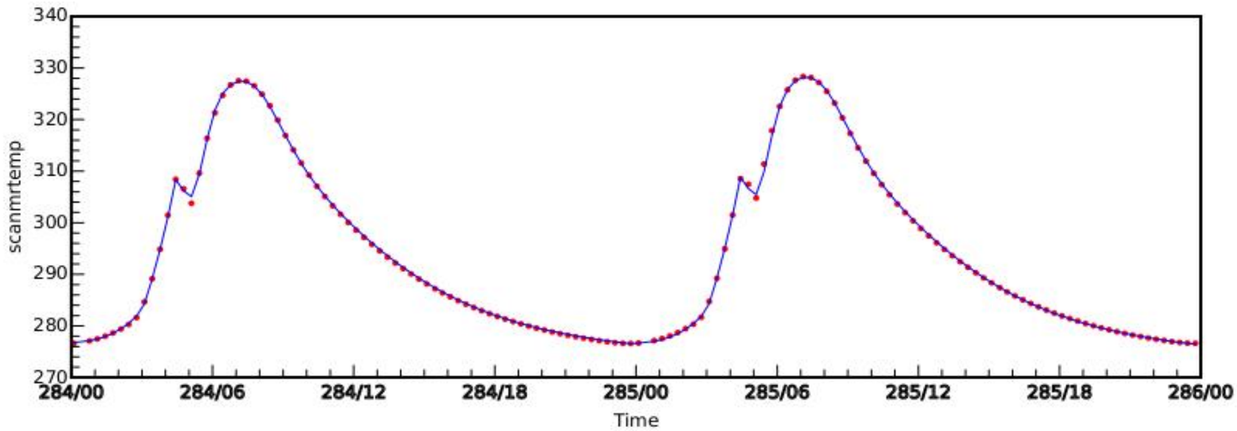

Figure 4 Initial training output of the scan mirror temperature for GOES 13 Sounder. The red dots represent the data points and blue line is the neural network output.

Figure 4 shows the data training results of the scan mirror temperature for GOES13 Sounder, which is the input used in the instrument data calibrations. The network structure for the temperature data is the same as that for the blackbody counts in Figure 3. Both data training results in Figures 3 and 4 show that the neural network provides excellent descriptions in both non-continuous region around the satellite midnight and in the continuous region. This shows how versatile the neural network is for datasets with different data patterns and scales.

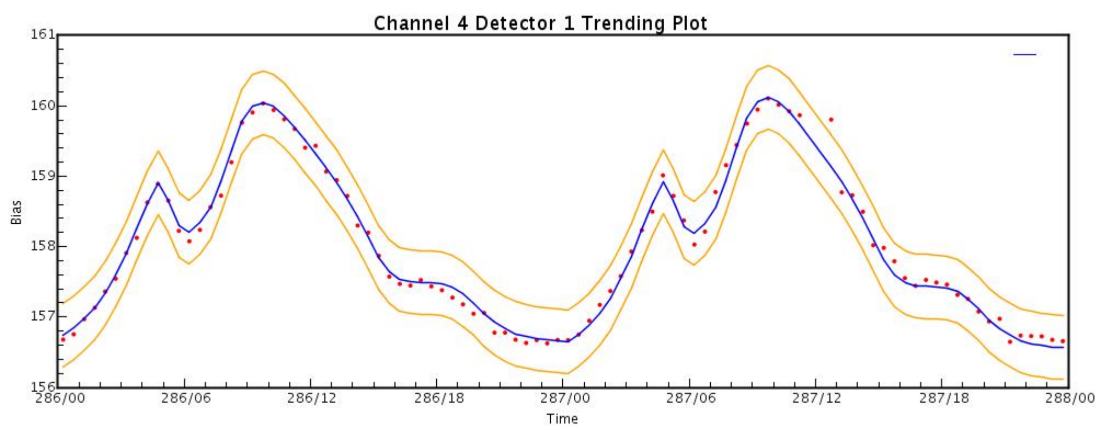

Figure 5 Output of data retraining of bias parameters for detector 1 in infrared channel 4.

The time dependent trend obtained from the data training enables the data monitoring, and Eq. 2 is evaluated for each data point. If the difference $|f(t_i) - d(t_i)|$ for a data point is larger than the value$N\sigma_e$, the data point is flagged as outlier that requires the engineer's attention. Figure 5 shows an example of outlier for the bias parameter in the infrared channel 4. The value of the data point around 13 hours on the day 287 is larger than the data bound being defined in Eq. 2, which could not be detected with static limits. The data can be monitored during the data training period after the time dependent trend has been obtained, and it could be done in real-time to calculate the difference between the predicted value based on the most recent time dependent trend and the data value.

5. Software Implementations

The machine learning system presented in this paper is implemented in Satellite Data Trending and Monitoring Toolkit (SDTMT). SDTMT is developed with Java and JavaFx technologies, which can be installed on most of Window and Linux platforms. Figure 6 shows the block diagram for SDTMT, in which the blocks in green represents the core SDTMT software. The red blocks in Figure 6 are project specific, which depends on the specific data structure and format. The functionalities of SDTMT are implemented to meet the operational needs for spacecraft engineering analysis, which include data trending, data monitoring, and historical data plots.

The data trending and monitoring blocks are two separate processes. The implementation of the trending engine uses the component approach, which is implemented as a component container that provides common interfaces for data trending, archive, and retrieval. Data trending algorithms are implemented as components in the container. Both ATLMA and machine learning algorithm are implemented trending in SDTMT.

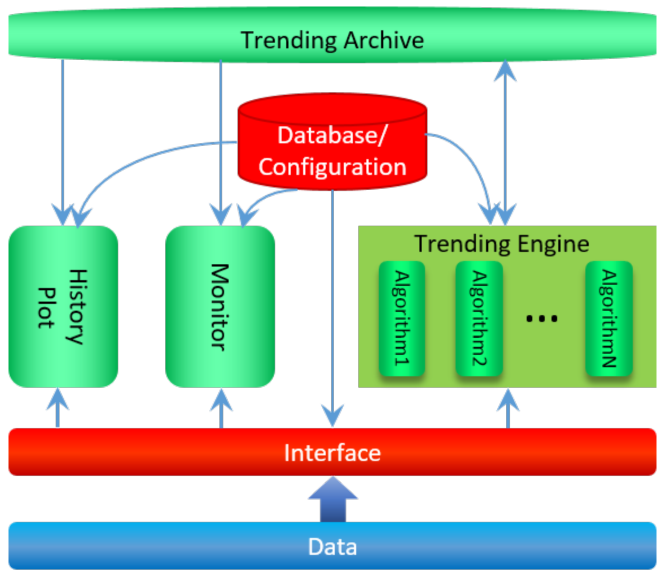

Figure 6 Block Diagram for Satellite Data Trending and Monitoring Toolkit

The operational concepts for SDTMT have two operation modes, which are the production mode and user interactive modes. The production mode runs the data trending, which generates the trending archive, trending plots, and outlier reports. The production mode runs once per day, and the outputs of the data training in the machine learning system are stored in a trending archive for the long-term storage. The user interactive mode provide a GUI for users to create and maintain the SDTMT configurations, perform manual data training or retraining, runs the real time data monitoring, and generates the historical plots. The trending data are displayed as the data plots. SDTMT implements a simple plot template called page used as plot definitions, and a page defines the global attributes, variable definitions and their plot attributes. Furthermore, SDTMT provides a dashboard for the real time monitoring of all variable defined in the database.

Both initial data training and retraining can be run in the production mode. The trending engine in SDTMT determines if a dataset is in initial training or retraining stage by searching the previous states in the trending archive. If the previous states exist, the data training performs the data retraining, otherwise, the initial training is invoked. The training output for datasets defined in the database can be examined through the user interactive mode, and a GUI component for data training has implemented for users to check the data training result and perform manual training if the training output is not satisfactory.

SDTMT defines a database schema, and the actual trending and monitoring database are being developed during the software deployment phase, which is specific to an environment being deployed. A SDTMT database defines the variable in a hierarchical structure for trending and monitoring, and it also defines the attributes needed for data training operation in a machine learning system. For example, the network structure parameter, *n*, is defined in the database for a specific variable based on its patterns. Users could adjust the parameter, *n,* during the software deployment phase to search for the network structure in the initial training.

The SDTMT defines a programming interface that converts the incoming data into the SDTMT internal format for trending and monitoring. SDTMT does not make any assumption on the format of datasets. The datasets that can be monitored by SDTMT are very diverse, which could include the spacecraft health and safety data that may have CCSDS standard format, the short term trend from TCS and other trending system, and the intermediate products in the instrument data processing process that may have special proprietary format. Thus, the interface processing is generally implemented during the system deployment phase.

6. Summary

The machine learning system presented here is a systematic and integrated approach to data trending and monitoring of the time dependent datasets. The trending of a dataset is a data training process in the machine learning, while the data monitoring is to determine if a data point is consistent with the time dependent trend generated from the data training processes. The feed-forward and back-propagation neural networks with two hidden layers are implemented as the machine learning algorithm, which have been proven to be very adaptive to different data patterns. The data training is the machine learning system implements two different data training strategies for the software deployment and normal operations. The data training strategy in the software deployment phase is generally a search process, while the data training during the normal operation uses the state variables from the previous training session as the input, and performs the fine tuning to the state variable to capture the seasonal or long term changes to dataset.

The machine learning system for data trending and monitoring provides a systematic approach for creating the situational awareness for a more proactive operations. It enables the real-time or near real-time monitoring of time dependent datasets that would have not been possible without an accurate time dependent trend. The system presented here is autonomous and adaptive. It requires little or no intervention from engineers during the normal operation phase. The seasonal and long-term changes to a dataset are captured through the data retraining. Thus, the machine learning system presented here presents a significant advance in spacecraft operations.

SDTMT will be deployed in GOES-R ground system for trending and monitoring the instrument calibration processes, and it potentially has wide range applications for trending and monitoring the spacecraft datasets as well as the datasets in instrument data processing process.

Reference

[1] Tom M. Mitchell, "Machine Learning", McGraw-Hill, (1997).

[2] Zhenping Li, C. Savkli, "Autonomic Computing for Spacecraft Ground Systems", Proceedings of Second IEEE International Conference on Space Mission Challenges for Information Technology, 2006. http://dx.doi.org/10.1109/SMC-IT.2006.21.

[3] Simon Haykin, "Neural Networks, A Comprehensive Foundation", Second Edition, Prentice Hall(1998).

[4] Zhenping Li, David Pogorzala, Ken Mitchell, J.P. Douglas, "Adaptive trending and limit monitoring algorithm for GOES-R ABI radiometric parameters" GSICS Quarterly Newsletter, Summer 2015 Issue, http://dx.doi.org/10.7289/V5XK8CHN#page9.

[5] G Cybenko, "Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient", (Technical Report). Department of Computer Science, Tuffs University, Medford, MA. (1988).

[6] C.M. Bishop, and C.M. Roach, "Fast Curve Fitting Using Neural Networks", Review of Scientific Instruments, 63.10 (1992), 4450, http://dx.doi.org/10.1063/1.1143696.

[7] H. Yu and B. M. Wilamowski, "Levenberg-Marquardt Training," in The Industrial Electronics Handbook, Vol. 5−Intelligent Systems, 2nd ed. (CRC Press, Boca Raton, 2011).

[8] M. Riedmiller, "Advanced Supervised Learning in Multi-Layer Perceptrons –From Backpropagation to Adaptive Learning Algorithms", Computer Standards & Interfaces, 16 (1992), 265, http://dx.doi.org/10.1016/0920-5489(94)90017-5.