A Health Index for Satellite System Based on Characteristics of Telemetry Data

Shun Katsube[1], Hironori Sahara[2]

[1,2]*Tokyo Metropolitan University, Department of Aeronautics and Astronautics, Hino, Tokyo, 191-0065, Japan*
katsube-shun@ed.tmu.ac.jp
sahara@tmu.ac.jpu

ABSTRACT

Since satellites are non-repairable systems, it is important to detect anomalies early to prevent failures. Data-driven approaches to anomaly detection, which have been actively studied in recent years, have problems such as low explainability and insufficient training data in initial operation. Thus, we propose a health index that is commonly available for all satellites. It is possible to share training data and examples of anomalies by monitoring health status with the same index across different satellites. In this study, we extended our health index defined for the satellite power system to the entire satellite system. Then, we applied the health index to the operational data of the Suzaku satellite and confirmed that the index is useful for anomaly detection.

1. INTRODUCTION

Anomaly detection is an important factor in satellite development because operators cannot repair launched satellites and must detect anomalies before severe failures occur. Thus, the requirements for anomaly detection in satellites are to detect anomalies as early as possible and to provide information that assists the operator in decision making, such as the anomaly type and the device that may be the cause of anomaly.

The most primitive anomaly detection method in satellites is limit checking, in which telemetry values are checked to determine whether they are within a predefined range. However, the complexity of current satellite systems and large number of onboard sensors make it difficult to set appropriate thresholds for all telemetry, and some anomalies occur within threshold values. Hence, expert systems (Takaki, Hashimoto, Honda, & Choki, 2006) and statistical model-based methods (Williams & Nayak, 1996) have been studied. These knowledge-driven methods help operators make decisions because of their high inference capability, but they require a lot of expert knowledge and human resources for operation and maintenance. In other words, operational limitations exist owing to the complexity of satellite systems. To remove these limitations, researchers have studied data-driven approach using unsupervised machine learning algorithms, and methods such as mixture of probabilistic principal components analyzers and categorical distributions (Yairi, Takeishi, Oda, Nakajima, Nishimura, & Takata, 2017), and Bayesian long short-term memory network (Chen, Pi, Wu, Zhao, Pan, & Zhang, 2021) have been applied to satellite telemetry. This approach is less demanding on developers than knowledge-driven approaches because it builds anomaly detection models inductively from historical data instead of expert knowledge. In addition, since a normal state model is learned from normal data, this approach judges all data that deviates from the model as anomalies and may inform the operator of unexpected anomalies or unknown findings. However, problems exist in that the model only presents the anomaly possibility and cannot explain the severity or causes of anomalies, and training data is insufficient in the early stages of satellite operation (Yairi, Fukushima, Liew, Sakai, & Yamaguchi, 2021).

To solve the problems of data-driven anomaly detection, we propose a health index for the satellite systems. This index is widely applicable to satellites and is a low-dimensional variable that varies corresponding to the unusual behavior of system. When a common standard health index is defined, it will be possible to share and compare training data for anomaly detection models among different satellites. Furthermore, the index behavior corresponds to the types of anomalies, which is expected to improve the explainability of anomaly detection model. In our previous work, we defined the health index for satellite power system and showed that the health index can represent anomalies such as voltage drops and thermal runaway (Katsube & Sahara, 2022).

In this study, we defined a new health index by introducing hyperbolic tangents in the preprocessing to improve the sensitivity of outlier detection and by adding features related to periodicity to make contextual anomalies detectable. We then applied it to telemetry data including subsystems other than the power system.

2. METHOD

2.1. Overview

To make the index applicable to various satellites, the index is defined based on characteristics that much telemetry data have in common. We assume the following three properties: telemetry data are time series data with periodicity corresponding to the orbital period of the satellite; telemetry data distribution has multiple clusters corresponding to the operational mode of the satellite; telemetry data under normal operation are within a certain value range. Thus, the data vector composed of telemetry data moves between each cluster within a certain period and value range in the data space. Anomalies in time series data are classified as non-contextual anomalies and especially temporal contextual anomalies (Ruff, Kauffmann, Vandermeulen, Montavon, Samek, Kloft, Dietterich, & Muller, 2021), so an anomaly state is when the data vector movement deviates from the normal value range and period.

Here, we focus on the distance y_k between the data vector $\boldsymbol{x}$ at time t and each cluster centroid $\boldsymbol{c}_k$, as shown in Figure 1 and denoted by Eq. (1).

$$y_k^{(t)} = d\left(\boldsymbol{x}^{(t)}, \boldsymbol{c}_k\right), \quad (k = 1, \dots, K), \tag{1}$$

where d is the distance function, k is the cluster number, and K is the number of operational modes. During normal operation, this distance has the same characteristics as telemetry data - clusters, periodicity, and value range - and when the data vector is in an abnormal position or does not follow a period, this distance also behaves unusually. Thus, we can use the distribution and time series models estimated from this distance respectively to monitor non-contextual and contextual anomalies.

To enable monitoring of contextual anomalies using only the distribution model, we introduce a new distance y_T between the data vector at time t and the data vector one cycle ago, represented by Eq. (2).

$$y_T^{(t)} = d\left(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t-T)}\right), \tag{2}$$

where T is the period of the telemetry data and k is the cluster number nearest to the data vector $\boldsymbol{x}^{(t)}$. This distance is distributed close to zero as long as the telemetry data maintains periodicity, and is an outlier when the periodicity is lost or changed.

Therefore, by monitoring the distribution of the vector whose elements are these distances, we can detect anomalies in the telemetry data. We define this vector $\boldsymbol{y}$ as a health index for satellite system as in Eq. (3).

$$\boldsymbol{y}^{(t)} = \left[y_1^{(t)}, \dots, y_K^{(t)}, y_T^{(t)}\right]^{\mathsf{T}} \tag{3}$$

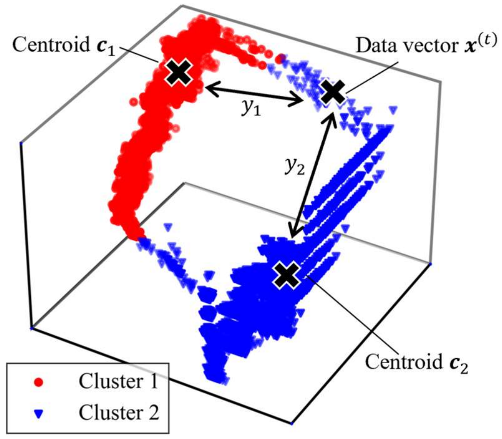

Figure 1. Distance between a data vector and two centroids.

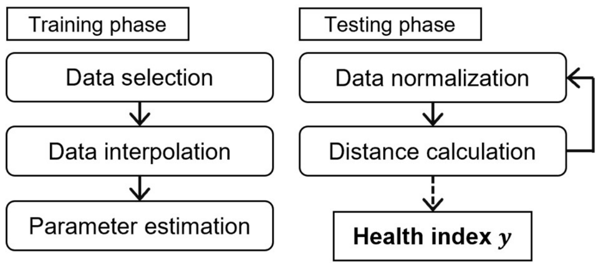

Figure 2. Flow of health index calculation.

The transformation to this index is represented by Eq. (4), which transforms a data vector $\boldsymbol{x}$ in a D-dimensional data space $\mathbb{R}^D$ consisting of D telemetry data into a data vector $\boldsymbol{y}$ in a $K+1$-dimensional data space $\mathbb{R}^{K+1}$.

$$f: \boldsymbol{x} \in \mathbb{R}^D \to \boldsymbol{y} \in \mathbb{R}^{K+1} \tag{4}$$

The process for calculating this index is shown in Figure 2 and described in the following sections.

2.2. Data Preprocessing

2.2.1. Data selection

In this study, we assume that telemetry is periodic, so we select periodically varying telemetry to create a dataset. Since the orbital period of the satellite Suzaku used for verification in section 3 is 96 minutes, only telemetry data whose autocorrelation peaks at lag 96 minutes were used. Specifically, among telemetry data with a sampling length of 60 and a sampling period of 4 minutes, we selected telemetry data whose autocorrelation peaked at lag 24 and was greater than 0.3.

Additionally, we simply assume that the operational modes are daytime and nighttime modes, and the number of operational modes in Eq. 1 is $K = 2$. Thus, the clustering in section 2.3.3 used one voltage, nine currents, and one temperature data related to power distributors and solar array paddles with daytime and nighttime modes.

2.2.2. Missing value interpolation

Because telemetry data has different sampling periods depending on the sensor, we resample to handle all telemetry data at each time step. In the validation of this study, the interval of all telemetry data was aligned to 32 seconds by down sampling with median.

2.3. Health index calculation

2.3.1. Data normalization

First, the telemetry data is normalized before calculating the distance as a health index. When data vectors have high dimensionality, outliers may be hidden by other normal data in the distance calculation. Hence, we transform the data vector $\boldsymbol{x}$ into $\boldsymbol{x}_{norm}$ combining normalization by the α th and $(100-\alpha)$ th percentile and the function $x^2 \tanh x$ to exaggerate the data vectors outside the normal value range, as in Eq. (5).

$$\boldsymbol{x}_{norm} = \left(\frac{2x-\boldsymbol{p}_\alpha-\boldsymbol{p}_{100-\alpha}}{\boldsymbol{p}_\alpha-\boldsymbol{p}_{100-\alpha}}\right)^2 \tanh\left(\frac{2x-\boldsymbol{p}_\alpha-\boldsymbol{p}_{100-\alpha}}{\boldsymbol{p}_\alpha-\boldsymbol{p}_{100-\alpha}}\right), \quad (5)$$

where $\boldsymbol{p}_\alpha$ is a vector whose components are the α th percentile of each data in the dataset. Here, determining α depends on how many outliers are expected to be included in the training dataset. In this study, the training dataset size was set to 162000 and $\alpha = 99.9$.

2.3.2. Distance calculation

Next, the distance is calculated from the normalized telemetry data. Since telemetry data are often strongly correlated, we use the Mahalanobis distance as the distance function d in Eq (1) and (2), which is more reasonable than the Euclidean distance when the data are correlated. Let $\boldsymbol{\Sigma}$ be the covariance matrix obtained from the dataset, the Mahalanobis distance is represented by Eq. (6).

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_1 - \boldsymbol{x}_2)} \quad (6)$$

2.3.3. Parameter estimation

The parameters required in the calculation process described above are the percentiles $\boldsymbol{p}_\alpha, \boldsymbol{p}_{100-\alpha}$, centroids $\boldsymbol{c}_1, \dots, \boldsymbol{c}_K$, and covariance matrix $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K$. First, the percentiles are computed from the training dataset, and then the centroid and covariance matrix are computed from the normalized training dataset by clustering using a Gaussian mixture model.

2.4. Application to anomaly detection

To validate the usefulness for system monitoring, we apply the health index to an anomaly detection method using a probability distribution model. Let $\hat{p}$ be the probability density function of the health index computed using kernel density estimation, and the anomaly score $a(\boldsymbol{y}^{(t)})$ is expressed by Eq. (7).

$$a(\boldsymbol{y}^{(t)}) = -\ln \hat{p}(\boldsymbol{y}^{(t)}) \quad (7)$$

3. Experiment

3.1. Dataset

We applied the health index to Suzaku satellite telemetry data, for which data is available on the web (ISAS/JAXA, 2016). Since the power generation has frequently dropped after 2011 in operation of Suzaku (Maeda, 2016), the telemetry data during 2007 was used as the normal dataset, and the telemetry data during the period including the voltage drop that occurred in January 2012 was used as the abnormal dataset (Table 1). Both datasets consist of 80 continuous telemetry data, which includes data for various physical units (Table 2).

An important point for the training dataset is that the percentiles for normalization can be properly computed. Because the datasets contained telemetry data with seasonal variations of two months, we set the training dataset for parameter estimation to the first 60 days within each dataset.

Table 1. List of datasets.

Dataset type	Period
Normal dataset	2007-01-01~2007-12-31
Abnormal dataset	2011-01-24~2012-01-24

Table 2. List of telemetry data.

Telemetry type	Unit	Number of data
Temperature	degC	51
Current	A	12
Voltage	V	4
Magnetic flux density	nT	4
Angle	deg, rad	3
Distance	km	3
Velocity	km/s	3

3.2. Normal operation

In the first experiment, we computed the health index with parameters estimated from the telemetry data for the period January 1, 2007 to March 1, 2007. We then computed the anomaly score with the probability density function estimated from the health index for the period March 2, 2007 to March 11, 2007. Those results are shown in Figure

3. Here, the health index and anomaly scores were processed with a moving median of window size 11, and the same is done in the following figures.

Figure 4 shows the distribution of the health index. In this figure, we can see that the health index has two clusters as well as the original telemetry data. And Figure 5 is a time series plot of the health index, which shows that the health index has the same periodicity as the original telemetry data. Thus, the health index retains the periodicity and multiple clusters that are properties of the original telemetry data.

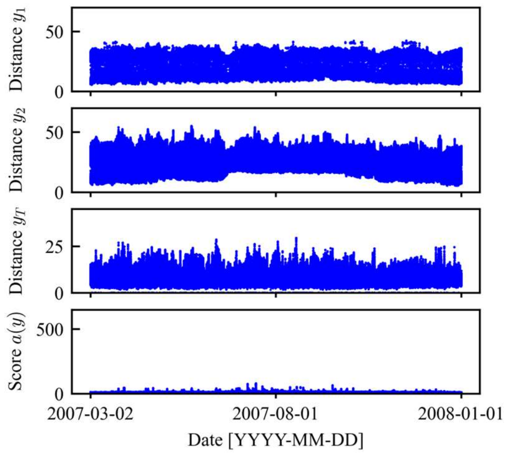

Figure 3. Health index and anomaly score in normal operation.

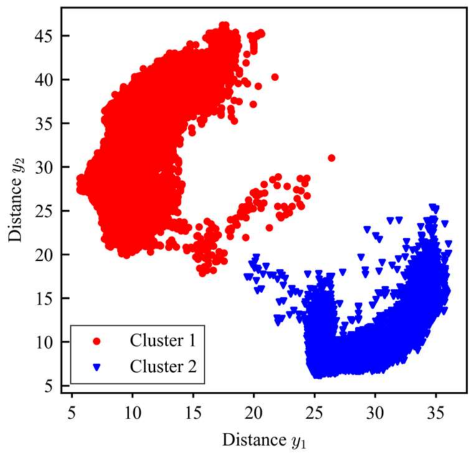

Figure 4. Distribution of health index elements y_1, y_2.

3.3. Abnormal operation: Voltage drop

In the second experiment, we computed the health index with parameters estimated from the telemetry data for the period January 24, 2011 to March 24, 2011. We then computed the anomaly score with the probability density function estimated from the health index for the period March 25, 2011 to April 3, 2011. Those results are shown in Figure 6.

Suzaku switched to SAFEHOLD mode on January 24, 2012, probably due to an anomaly in the power system equipment (ISAS/JAXA, 2015). In the gray area of Figure 6, there are two increases in the anomaly score, one occurring just before switching to SAFEHOLD mode and the other before that. The former increase was caused by a drop in battery voltage, which indicates that the health index can represent an anomaly. The latter increase was likely caused by a rise in some temperature data, but it is unclear if this is a sign of the anomaly, another anomaly, or a rare event. In addition, since the distances y_1, y_2, y_T increased in Figure 6, the anomaly occurring in Suzaku is considered to be a non-contextual anomaly, which means that some data deviates from the normal value range. On the other hand, only the distance y_T is expected to increase in the case of contextual anomalies. This result indicates the potential for estimating the type of anomaly from the behavior of the health index, because the actual anomaly is a voltage drop, which is a non-contextual anomaly. In actual operation, we consider that anomalies are detected based on a threshold determined from the variance and then the type of anomaly is identified from the components of the index.

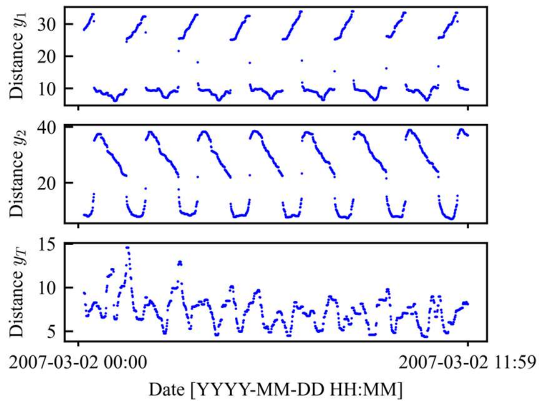

Figure 5. Time series plot of health index.

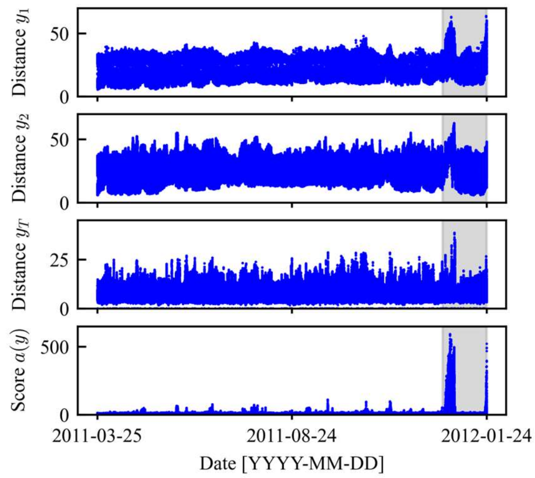

Figure 6. Health index and anomaly score in abnormal operation.

Simple anomalies in this case can be found manually, but complex anomalies, such as those caused by the involvement of multiple data, are difficult to detect. We consider that including data relationships as a component of the index will allow for more enhanced detection and classification of anomalies.

4. CONCLUSION

In this study, we defined the health index for our satellite system and applied it to the operational data of the Suzaku satellite. The health index is based on the periodicity and multiple clusters of the telemetry data and retains these properties. As a result of our experiments, we found that the health index can represent anomalies actually occurring on Suzaku and is useful for anomaly detection. We also show the potential of monitoring the behavior of health index to estimate the type of anomaly.

In the future, we will extend the assumptions about the properties on which health index is based. For example, increasing the number of operational modes and considering non-periodic and status data. In addition, we will share the index among different satellites and apply it to various anomaly detection methods to verify its usefulness.

ACKNOWLEDGEMENT

This work used data obtained from the Data ARchives and Transmission System (DARTS) operated by the Institute of Space and Astronautical Science (ISAS), Japan Aerospace Exploration Agency (JAXA) and the Scientific Satellite Operation and Data Archive Unit (C-SODA).

This work was supported by JST SPRING, Grant Number JPMJSP2156.

REFERENCES

Chen, J., Pi, D., Wu, Z., Zhao, X., Pan, Y., & Zhang, Q. (2021). Imbalanced Satellite Telemetry Data Anomaly Detection Model Based on Bayesian LSTM. *Acta Astronautica*. vol. 180, pp. 232-242.

ISAS/JAXA (2015). Log of important operation events and happenings. http://www.astro.isas.jaxa.jp/suzaku/log/operation/.

ISAS/JAXA (2016). Data archive and transfer system (DARTS). https://darts.isas.jaxa.jp/astro/suzaku/.

Katsube, S., & Sahara, H. (2022). Proposal of an Index of Satellite Health for Anomaly Detection. *73rd International Astronautical Congress*. September 18-22, Paris, France.

Maeda, Y. (2016). A Story of the Operation of the Suzaku Battery and the Solar Array Paddle, *The Astronomical Herald*, vol. 109, pp. 14-20.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Muller, K.-R. (2021). A Unifying Review of Deep and Shallow Anomaly Detection, *Proceedings of the IEEE*, vol. 109, pp. 756-795.

Takaki, R., Hashimoto, M., Honda, H., & Choki, A. (2006). ISACS-DOC: Automatic Monitoring and Diagnostic System for Scientific Satellite, *Japanese Society for Artificial Intelligence*, vol. 21, pp. 20-25.

Williams, B. C. & Nayak, P. P. (1996). A Model-based Approach to Reactive Self-configuring Systems, *Proceedings of 13th National Conference on Artificial Intelligence*, vol. 2, pp. 971-978.

Yairi, T., Takeishi, N., Oda, T., Nakajima, Y., Nishimura, N., & Takata, N. (2017). A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Probabilistic Clustering and Dimensionality Reduction. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, pp. 1384-1401.

Yairi, T., Fukushima, Y., Liew, C. F., Sakai, Y., & Yamaguchi, Y. (2021). A Data-Driven Approach to Anomaly Detection and Health Monitoring for Artificial Satellites. *Advances in Condition Monitoring and Structural Health Monitoring*, pp. 129-141.