International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

Performance Degradation Analysis Method using Big Data

M. Bhargavi,
Assistant Professor,
Department of CSE,
Vidyanikethan Engineering College,
Tirupati,

S. Vijayalakshmi,
Assistant Professor,
Sree Vidyanikethan Engineering College,
Tirupati,

C. Venkata Lokesh,
Student Scholar,
IV CSE, Department of CSE,
Sree Vidyanikethan Engineering College,
Tirupati,

Abstract. **Satellites have features of high control integration, various working modes, and complex telemetry big data, which make it difficult to evaluate their performance degradation. In this paper, a novel data mining analysis method is proposed to analyze the satellite's telemetry big data, in which sample entropy is calculated to characterize states and support vector data description is utilized to analyze the satellite performance degradation process. The experimental results show that our proposed method could generally describe the performance degradation process of satellites. Meanwhile, it also provides an important approach for the ground-station-monitor to analyze the performance of satellites.**

Keywords: Performance degradation; Telemetry big data; Sample entropy; Support vector data description

1. INTRODUCTION

With more and more satellites being sent into space these years, the ground in-orbit managements have to handle such challenges as satellite's high control precision, various working modes, and high complexity. As advanced technologies and new materials are utilized in satellites, the sudden failure is not the primary failure mode for most satellite failures, which is replaced by performance degradation. The theory of analyzing satellite performance degradation only focuses on the overall performance of equipment, regardless of failure modes, which is different from analyzing sudden failures.

In 2001, the University of Wisconsin and the University of Michigan, together with other 40 industry partners, were united to establish the Intelligent Maintenance Systems (IMS) research center under the U.S. National Science Foundation. After then, many methods of performance degradation assessment have been proposed, such as the pattern discrimination model (PDM) based on a cerebellar model articulation controller (CMAC) neural network [1], self-organizing map (SOM) and back propagation neural network methods [2], hidden Markov model (HMM) and hidden semi-Markov model (HSMM) [3], etc. However, these methods are deficient in some aspects. For example, the results of CMAC assessment method are greatly influenced by parameter setting, and the assessment results of the SOM, neural network method and hidden Markov model cannot directly reflect degradation degree. In order to accommodate the characteristics of assessment for different key components, the analysis theory of performance degradation has been developed from single degradation variable to a more diverse practical direction. Although some new theories and methods have emerged, the researches on the performance degradation of satellite are still limited. M Tafazoli [4] studied in-orbit failures for more than 130 different spacecraft and revealed that the spacecraft are vulnerable to failures occurring in key components. MA W [5] analyzed the space radiation environment of thermal coatings and proposed degradation models for the optical properties of thermal coatings. However, these methods mainly focus on failure data and also require relevant experience.

The conventional analysis methods for satellite performance degradation have some shortcomings such as experimental difficulties and high cost. According to expert knowledge, large amounts of telemetry big data are generated during the in-orbit operation and monitoring process. Satellites telemetry big data contain monitoring information, abnormal states, space environment, and others, which reflect the operational status and payload of satellites. A novel analysis method for satellite performance degradation with telemetry big data is proposed in this paper. This method uses data mining techniques and provides a quantitative description for satellite performance degradation process. Furthermore, it also can be extended to apply to failure prediction.

2. RELATED CONCEPTS

2.1. Sample Entropy

The sample entropy [6] (SamEn) is an improved algorithm of approximate entropy (ApEn) proposed by Pincus [7]. The advanced algorithm is able to quantify the complexity rate of a nonlinear time series.

For a data series $X(N) = x(1), x(2), ... x(n)$, where N is the length of the series, two parameters are defined: m is the embedded dimension of the vector to be formed and r is the threshold that serves as a noise filter. The steps to calculate SamEn are shown as follows:

1) $\mathrm{N}-m+1$ patterns (vectors) are generated, and each pattern owns m dimensions. The pattern is represented as following:

$$X^m(i)=\left[x(i),x(i+1)\ldots x(i+m-1)\right] \quad i=1,\cdots,N\text{-}m\text{+}1 \tag{1}$$

2) The distance, $d\left[X^m(i),X^m(j)\right]$ between each two patterns can be computed by using Eq. (2).

$$d\left[X^m(i),X^m(j)\right]=\max\left[\left|x(i+k)-x(j+k)\right|\right] \quad k=0,m\text{ - }1 \tag{2}$$

3) For each pattern $X^m(i)$, the number of matching pattern, $N^m(i)$, i.e., number of $d\left[X^m(i),X^m(j)\right]\le r$ is achieved. $C_r^m(i)=N^m(i)/(N-m+1)$ is the probability that pattern $X^m(j)$ matches $X^m(i)$. And the matching probability of two sequences with m points can be achieved by using Eq.(3):

$$\Phi^m(r)=\frac{1}{N-m+1}\sum_{i=1}^{N-m+1}C_r^m(i) \tag{3}$$

4) When the dimension expands to $m+1$, steps 1-3 are repeated to find out $\Phi^{m+1}(r)$. The theoretical value of the SamEn is defined as follows:

$$\mathrm{SamEn}(m,r)=\lim_{N\to\infty}\left[\ln\left(\Phi^m(r)-\Phi^{m+1}(r)\right)\right] \tag{4}$$

For a finite length of data points N, the estimated value of the SamEn is given by using Eq.(5):

$$\mathrm{SamEn}(m,r,N)=\ln\left[\Phi^m(r)-\Phi^{m+1}(r)\right] \tag{5}$$

Experiments conducted by Pincus [8] indicate that a reasonable statistical character can be achieved when $m=2$, $r=(0.1\sim0.25)\cdot std(X)$, where $std(X)$ denotes the standard deviation of $X=\{x_1,x_2,\cdots x_N\}$.

Compared with the general nonlinear dynamics, SamEn has more advantages, such as immunity to noise and inference as well as independence from the length of time series. SamEn has been widely used in physiological signal processing because of its excellent characteristics [8, 9]. In consideration of these characteristics, SamEn is a promising method in describing the performance features for a large amount of telemetry big data. The SamEn is also used in this paper to extract satellite performance features.

2.2. Support Vector Data Description

Support Vector Data Description [10] (SVDD) is inspired by the Support Vector Classifier. The method is robust against outliers in the training set and is capable of tightening the description by using negative examples.

A hypersphere that contains all or most samples of the target class is defined as $X=\{x_1,x_2,\cdots x_n\}$. The hypersphere is bounded by the core of the hypersphere a and radius R. If the hypersphere covers all the training samples of target class, the classification is established by the empirical error which is equal to zero, and the structural error is defined as follows:

$$\varepsilon(a,R)=R^2. \tag{6}$$

As the distance from x_i to the core a should not be larger than radius R for all the samples of the target class X, the constraint of the minimization problem can be described as Eq. (7):

$$\|x_i\text{-}a\|^2\le R^2 \tag{7}$$

To account for the possibility of outliers in the training set, the distance between x_i and the core a should not be strictly smaller than R, but larger distances should be penalized. Therefore, slack variable ξ_i is brought in, and the minimization problem is transformed into

$$\left.\begin{aligned}&\min\ \varepsilon(R,a,\xi)=R^2+C\sum_{i=1}^{N}\xi_i\\&s.t\ \|x_i\text{-}a\|\le R^2+\xi_i\\&\xi_i\ge0\ (i=1,2,\cdots,N)\end{aligned}\right\} \tag{8}$$

The penalty factor C makes a trade-off between the volume and the errors. The minimization problem in Eq.(8) can be calculated by using Eq.(9).

$$\begin{aligned}L(R,a,\alpha_i,\xi_i)=&R^2+C\sum_i\xi_i\\&-\sum_i\alpha_i\left\{R^2+\xi^2-\left(x_i-2ax_i+a^2\right)\right\}\\&-\sum_i\gamma_i\xi_i\quad(\alpha_i\ge0,\gamma_i\ge0)\end{aligned} \tag{9}$$

In Eq. (9), α_i and γ_i are the Lagrange multipliers. L should be minimized with respect to R, a, and ξ_i and maximized with respect to α_i and γ_i. Respectively taking their partial derivatives equal to zero, and then get the following constraint Eq.(10):

$$\left.\begin{aligned}&\sum_i\alpha_i=1\\&a=\frac{\sum_i\alpha_ix_i}{\sum_i\alpha_i}=\sum_i\alpha_ix_i\\&C-\alpha_i-\gamma_i=0\quad\forall i\end{aligned}\right\} \tag{10}$$

Substituting (10) into (9),we obtain $\max L$:

$$\max L=\sum_{i=1}\alpha_i(x_i\cdot x_i)-\sum_{i,j}\alpha_i\alpha_j(x_i\cdot x_j) \tag{11}$$

According to the theory proposed by Vapnik [11], the Kernel trick can be adopted to take the place of dot product. Using the kernel function enables SVDD to handle the mapping of low-dimensional original space to high-dimensional feature space without dimensional disaster. Any function satisfies the Mercer's condition can be regarded as the kernel function, and RBF [12] is used as the kernel function in this work:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 09, September-2016

$$K_G(x,y,\sigma) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right) \quad (12)$$

The optimization problem described by Eq. (11) can be further transformed into the following explicit form:

$$\max L = 1 - \sum_{i,j} \alpha_i \alpha_j K_G(x_i, x_j, \sigma) \quad (13)$$

Equation (13) shows that the core of the hypersphere is a linear combination of the objects. Only objects x_i with $\alpha_i \geq 0$ are needed in the description. Therefore, these objects are called the support vectors of the description (SVs). To test an object z, the distance to the core of the hypersphere and the radius R are respectively calculated by Eq. (14).

$$d = \|z\text{-}a\| = K_G(z,z) \text{-} 2\sum_i \alpha_i K_G(z, x_i) + \sum_{i,j} \alpha_i \alpha_j K_G(x_i, x_j) \quad (14)$$

$$R^2 = \|x_{sv}\text{-}a\|^2 = 1\text{-}2\sum_i \alpha_i K_G(x_i, x_{sv}) + \sum_{i,j} \alpha_i \alpha_j K_G(x_i, x_j) \quad (15)$$

The test object z is accepted when this distance is not greater than the radius (i.e. $d \leq R$).

SVDD has the advantage of requiring only one category as the learning sample, whereas the degradation analysis itself plays down the distinction between specific patterns. The sample points of health status are extracted as the learning samples. Therefore, the process of moving away from the health status with time for the testing samples can be regarded as the degradation process. Only the core of the hypersphere is used to detect the target class of testing sample. And the core can be determined by a few support vectors. Moreover, the satisfactory computational speed of SVDD to classify the testing samples makes SVDD a promising alternative method for analyzing satellite performance degradation.

3. METHOD TO ANALYZE THE PERFORMANCE DEGRADATION OF SATELLITE

3.1. Definition description

Definition 1 (Performance Eigenvector)

The SamEn of a time period is taken as its performance feature. And the vector composed of the performance features of parameters within the same time period is called performance eigenvector.

In this study, parameters are not limited to those of the objective equipment, but they also contain a number of closely related equipment parameters. As parameters are relative to specialized knowledge, their selections are conducted based on the domain and expert knowledge.

Definition 2 (Health Model)

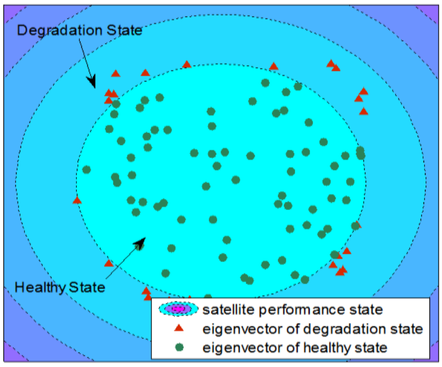
With SVDD method, the model obtained by training the performance eigenvector of satellite in the healthy status is called health model (*model*).

According to the theory of SVDD, the model described in definition 2 is composed of the support vectors of healthy state vector (*model.SV*), corresponding coefficients (*model.σ*),number of support vectors (*model.len*), hypersphere bounded by the core (*model.a*) and the radius (*model.R*)

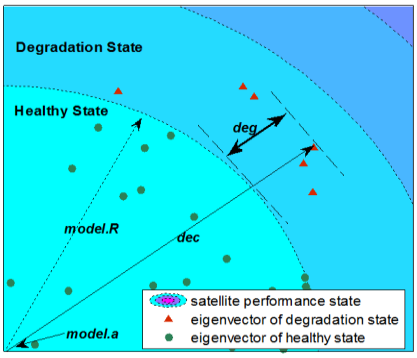
Definition 3 (Performance Degradation Degree)

Here, *dec* denotes the distance between the performance eigenvector of satellite and the core of hypersphere. The performance degradation degree *deg* is defined by the difference between *dec* and the radius of hypersphere *model.R*, that is, *deg* = *dec* – *model.R* (in Figure 1).

It means that performance degradation process of the objective equipment may occurs when the value of *deg* is larger than 0. When the value increases monotonously, the performance degradation process of the objective equipment increases accordingly. As the degree cannot be negative, set *deg* = 0 when *dec* – *model.R* <*0*.



a) Performance states and eigenvectors



b) Performance degradation degree

Fig. 1. Principle of the performance degradation degree

Figure 1 shows the principle of performance degradation degree. However, the model cannot contain all the health status features of the satellite for the operating mode of satellite is complex, and the training sets in healthy status of each operating mode are limited. A satellite may remain in the healthy status under other operating modes, especially when *deg* is positive. Therefore, Definition 3 is appropriate for the parameters less affected by the operating mode of satellite.

3.2. Framework description

Figure 2 shows the overall framework of the analysis for satellite performance degradation presented in this study, which has four main steps.

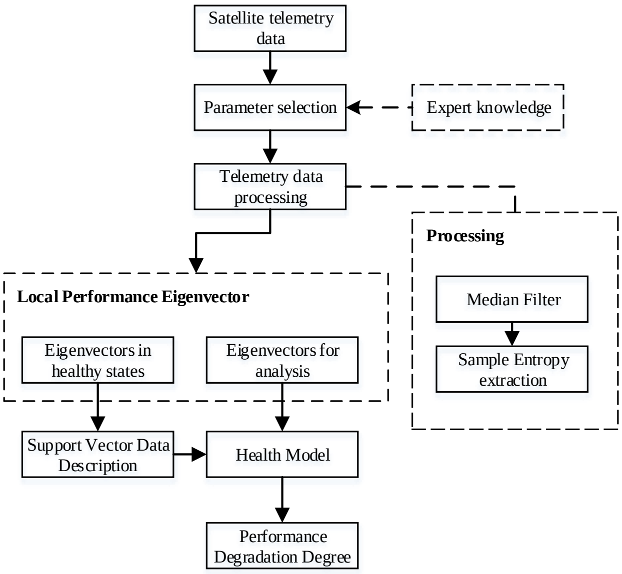


Fig. 2. Framework of the satellite performance degradation analysis

Step 1. Selectparameters of the satellite according to expert knowledge. Then, median filter method is used to reduce the noise in satellite telemetry big data so as to generate a new clean dataset.

Step 2. Extract the performance features from the selected parameters through Step 1 according to Definition 1. And compose the final set of the performance eigenvectors.

Step 3. Select the performance eigenvectors in the healthy status as the training set, and build a health model with SVDD method.

Step 4. To measure the degradation status of the new performance eigenvector, calculate the performance degradation degrees according to Definition 3 and the results of the model obtained in Step 3.

Considering the features of satellite telemetry big data, the median filter method in Step 1 is used to reduce noise in the data.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The telemetry big data of one satellite is used as experimental data, which recorded from 2011-05-01 00:00:00.0 to 2011-12-29 18:16:59.987, 14 million data frames that contain several failures and performance degradation information. In our experiments, seven important parameters in this dataset are selected by expert knowledge.

The telemetry big data is stored in Oracle 11g, and the algorithms are coded by Java. The operating system used is Windows Server 2008 R2 Standard with the Intel (R) Xeon (R) Eight-core E5606 processor with 8 G RAM.

4.1. Telemetry big data processing

The experimental dataset is processed as the following steps:

(1) The outliers caused by decoding or other errors are removed according to the ranges of the seven parameters. And further, the median filter method in every 30s is used to reduce the noise in the dataset. Finally, a new dataset is achieved.

(2) The values of time series are normalized into the range [-1,1] for each parameter and each time series are equally divided into 800 groups. The performance features of each group are extracted by Definition 1. Finally, seven performance feature sequences are obtained with a length of 800. The performance eigenvector is composed of the features of seven parameters in the group with same number.

4.2. Modeling and degradation analysis

(1) Performance eigenvector under healthy status are selected as the training data, SVDD method is used by setting σ=1 in this experiment, and then the health model of satellite is established.

(2) The remaining dataset is used as test data to verify the obtained health model, and the degradation degree is calculated according to Definition 3. Figure 3 shows the final results.

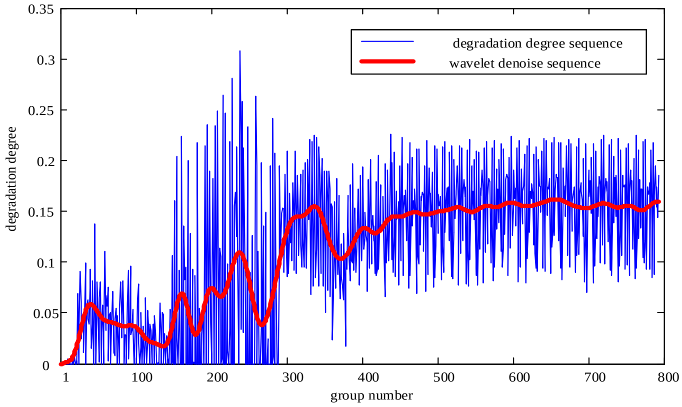


Fig. 3. Degradation degree

The degradation degrees are unsteady, and the curve is not smooth but fluctuant. This is mainly due to the recognition accuracy of SVDD and cyclical factors of original data that does not affect the overall reaction on the degradation process of satellite. In order to reduce the interference of these factors, a relative algorithm [13] is employed and the wavelet denoising sequence is obtained as Figure 3 shows. Overall, the average degradation degree presents an increasing trend. Given the long period, the accidental factors cannot influence the degradation degree all the time. Therefore, we conclude that the satellite has entered the performance degradation state based on Definition 3.

Aerospace experts confirm that two major failures of satellite did occur from late July to late August (between the 246th group and 370th group) for unknown reasons, and these two failures are corresponding to the two peaks nearby. That proves the correctness of our proposed definition, especially explaining the degradation peak and the high degradation degree level after the peak. In conclusion, the proposed method can efficiently describe the performance degradation process of satellite.

5. CONCLUSIONS

A method for satellite performance degradation with telemetry big data is proposed in this paper while studies for solving this problem are limited. The experimental analysis shows that the proposed method can extract effective state information from the parameters and provide a quantitative description for satellite performance degradation. Moreover, the analysis on the performance degradation of satellite with telemetry big data has a significant meaning in in-orbit research and management for satellites.

In our study, the definitions may have some limitations; for example, the degradation degree of the experiment is unstable but fluctuant. The sample entropy algorithm may take much time to trim redundant parameters in massive data, which will be improved in our future work.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (Grant No. U1433116).

REFERENCES

[1] Lee J. Measurement of machine performance degradation using a neural network model[J]. Computers in Industry, 1996, 30(3): 193-209.

[2] Huang R, Xi L, Li X, et al. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods[J]. Mechanical Systems and Signal Processing, 2007, 21(1): 193-207.

[3] Si X S, Wang W, Hu C H, et al. Remaining useful life estimation–A review on the statistical data driven approaches[J]. European Journal of Operational Research, 2011, 213(1): 1-14.

[4] Tafazoli M. A study of on-orbit spacecraft failures[J]. Acta Astronautica, 2009, 64(2): 195-205.

[5] MA W, XUAN Y, HAN Y, et al. Degradation Performance of Long-life Satellite Thermal Coating and Its Influence on Thermal Character [J]. Journal of Astronautics, 2010, 2: 043.

[6] Widodo A, Shim M C, Caesarendra W, et al. Intelligent prognostics for battery health monitoring based on sample entropy[J]. Expert Systems with Applications, 2011, 38(9): 11763-11769.

[7] Pincus S M. Assessing serial irregularity and its implications for health[J]. Annals of the New York Academy of Sciences, 2001, 954(1): 245-267.

[8] Alcaraz R, Rieta J J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms[J]. Biomedical Signal Processing and Control, 2010, 5(1): 1-14.

[9] Yang A C, Huang C C, Yeh H L, et al. Complexity of spontaneous BOLD activity in default mode network is correlated with cognitive function in normal male elderly: a multiscale entropy analysis[J]. Neurobiology of aging, 2013, 34(2): 428-438.

[10] Weinshall D, Zweig A, Hermansky H, et al. Beyond novelty detection: Incongruent events, when general and specific classifiers disagree[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 1886-1901.

[11] Yu H, Xie T, Paszczynski S, et al. Advantages of radial basis function networks for dynamic system design[J]. IEEE Transactions on Industrial Electronics, 2011, 58(12): 5438-5450.

[12] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.

[13] Shao Q, Zhang X, Qi X, et al. Optical wavelet de-noising applied in multi-span nonlinear fiber links[J]. Optics Communications, 2010, 283(7): 1261-1267.