A data-driven prognostics approach for RUL based on principle component and instance learning

Li Yongxiang, Shi Jianming, Wang Gong, Liu Xiaodong
Technology and Engineering Center for Space Utilization
Chinese Academy of Science
Beijing, China
liyongxiang@csu.ac.cn

***Abstract*—The research of Remaining Useful Life (RUL) estimation is one of the most common tasks of Prognostics and Health Management (PHM). This paper presents a data-driven approach for estimating RUL using principle component and instance learning. The approach is especially suitable for situations in which abundant run-to-failure (RtF) data are available. Firstly, the principal component analysis (PCA) is used to find the low-dimensional principal components (PCs) from the statistical features of the measured signals. Then, the health indicators (HI) can be obtained by using weighted Euclid distance (WED), and regressed by the data-driven methods or model-based methods. Finally, the method based on instance learning is employed to estimate the RUL of the machine under operation. The performance of the prognostics approach introduced in this paper is demonstrated by using turbofan engine degradation simulation data set, which is supplied by NASA Ames.**

Keywords-RUL, PCA, health indicators, failure threshold, instance learning

I. INTRODUCTION

PHM is a system engineering discipline focusing on detection, prediction, and management of the health and status of complex engineered system, and the research of RUL estimation is one of the most common tasks of PHM. The methods to calculate the RUL are generally divided into model-based methods and data-driven methods. The most commonly used model-based methods is physics of failure method, which is usually applied at material level or component level, and derived from wear or failure mechanism [1,2].

Data-driven methods are becoming popular due to their intuitive nature, fast developmental cycle and data storage and processing technologies. The commonly used data-driven methods include linear/non-linear regression, ARMA [3], Artificial Neural Networks, Fuzzy Logic system [4] and so on.

In this paper, a data-driven approach for estimating RUL is introduced. The approach, divided into training process and testing process, is especially suitable for situations in which abundant RtF data are available. In the training process, the PCA is used to reduce dimensions, WED is used to get the HI, and the data-driven methods are used for regression. In the testing process, the method based on instance learning is employed to estimate the RUL of the machine under operation. The advantages of this method proposed in this paper are as follows, firstly, it is suit for the situation that some of the important information, such as fault modes, operational conditions and so on, is missing due to lack of knowledge and incomplete measurements, secondly, we need not to consider the represent of the parameters, the only point we concerned is whether the data has a trend over time.

The rest of this paper will be organized as follows. Section 2 will provide a detailed description of prognostics approach. Firstly, the critical components should be identified and the corresponding physical parameters would be selected. Secondly, after outlier removal and normalizing, the dimensions of the data set will be reduced with the help of PCA, then the HI and failure threshold (FT) can be obtained by weighted Euclid distance and the curve of HI can be regressed by data-driven method. At last, the method based on instance learning will be employed to estimate the RUL of the machine under operation. Section 3 will provides a case, turbofan engine degradation simulation data set, will be applications and the results will be depicted. Section 4 concludes the paper.

II. PROGNOSTICS APPROACH FOR RUL BASED ON PRINCIPLE COMPONENT AND INSTANCE LEARNING

The RUL prognostics approach is divided into two parts: training process and testing process (Fig. 1). Training process contains training data (offline data) preparation, outliers removal, normalization, dimension reduction, HI and failure threshold (FT). While testing process contains testing data (online data) preparation, data processing, instance learning and RUL obtained. In testing process, testing data processing contains normalization and dimension reduction.

978-1-5090-0382-2/16/$31.00 ©2016 IEEE

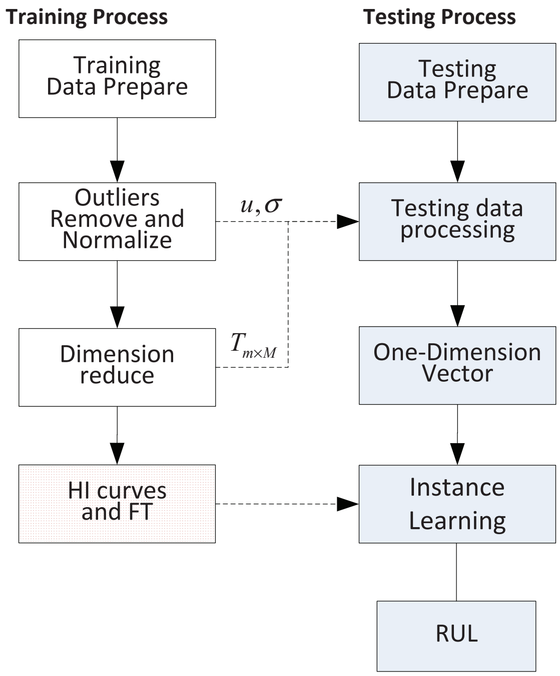

Figure 1. The data processing of prognostics approach

The u , σ and T in Fig. 1 are the mean, standard deviation and the transition matrix of the training data, which will be used in the testing process.

A. Training Process

The training data processing introduced here is just like the Fig. 1 shows. The detailed steps are as follows.

1) Data preparation

The data can be got from different sensors. For example, the lithium battery data set would contain charge and discharge voltage/current, temperature and impedance. In paper [5], five steps are divided to prepare the data: identification of critical components, selection of physical parameters, selection of monitoring sensors, data acquisition and data pre-processing. The method how to get the data set would not be discussed in this paper.

2) Outliers Removal

As the PCA is sensitive to the outliers, so before the dimension reduction, the outliers should be removed firstly. The method here to remove outliers contains outliers' detection and replacement.

Firstly, Equation (1) is introduced to get the local mean of the train data. Considering train data is $X=\{x_1,x_2,...,x_n\}$, and p here is constant, and the value is according to the length of the data.

$$\bar{x}_i = \begin{cases} \dfrac{1}{2p+1}\sum\limits_{j=i-p}^{i+p} x_j \,, p<i<n-p \\ \dfrac{1}{i+p}\sum\limits_{j=1}^{i+p} x_j \,, i\le p \\ \dfrac{1}{n-i+p+1}\sum\limits_{j=i-p}^{n} x_j \,, i\ge n-p \end{cases} \qquad (1)$$

Where, n is the length of training data, and $\bar{x}_i$ is local mean value of x_i .

Secondly, the absolute deviation $\Delta X=\{\Delta x_1,\Delta x_2,...,\Delta x_n\}$ and mean absolute deviation (MAD) of X and $\bar{X}$ can be got from (2) and (3).

$$\Delta x_i = | x_i - \bar{x}_i | \qquad (2)$$

$$\Delta x_{MAD} = \frac{1}{n}\sum_{i=1}^{n}\Delta x_i \qquad (3)$$

Thirdly, compare Δx_i and Δx_{MAD} as (4), and determine the value.

$$x_i = \begin{cases} \bar{x}_i ,\ if\ \Delta x_i > k\cdot\Delta x_{MAD}, \\ x_i ,\ else \end{cases} \qquad (4)$$

3) Normalization

To make the results equitable, the normalization is used. There are two normalization methods, Min-Max scaling and Z-score standardization, are discussed here.

Suppose that the measured signal (train data) is $X=\{x_1,x_2,...,x_N\}$, $x_{i\_\min}$ and $x_{i\_\max}$ are the min and max value of the vector x_i (denoted as the vector $\{x_{1i},x_{2i},...,x_{mi}\}$), $\bar{x}_i$ and σ_i are the mean and standard deviation of x_i . The comparison of the two normalizations is show as Table. 1.

TABLE I. THE COMPARISON OF MIN-MAX SCALING AND Z-SCORE STANDARDIZATION

	Min-Max scaling	**Z-score standardization[6]**
Equation	$x_{ij\_norm} = \dfrac{x_{ij}-x_{i\_\min}}{x_{i\_\max}-x_{i\_\min}}$	$x_{ij\_norm} = z(x_{ij}) = \dfrac{x_{ij}-\bar{x}_i}{\sigma_i}$
Scaling	[0, 1]	M=0，σ=1
Advantages	Easy and simple.	Suit for the calculation concerning the variance and covariance.
Disadvantages	The answer is inevitably influenced by the variance and covariance.	The data obey normal distribution.

Considering the influence of the covariance (see Appendix B), we choose the Z-score standardization to normalize the train data.

4) Dimension reduction

Suppose $Y_{n\times m} = \{y_{ij} \mid i = 1,2,..,n; j = 1,2,..,m\}$,and $Y_{n\times m}$ is the train data after normalizing. Then the steps using PCA are as follows.

a) Calculate the Correlation Matrix (CM).

The CM is $R_{m\times m} = \{r_{kj} \mid k = 1,2,..,m, j = 1,2,..,m\}$, and $r_{kj} = \dfrac{Cov(y_k, y_j)}{\sigma_k \cdot \sigma_j}$.

b) Eigenvalues and eigenvectors

Then, the eigenvalues $\{\lambda_1,...,\lambda_m; \lambda_1 > ... > \lambda_m\}$ and eigenvectors $\{v_1,...,v_m\}$ can be obtained from $|\lambda I - R_{m\times m}| = 0$.

c) Transition matrix

The transition matrix is $T_{m\times M} = \{v_1, v_2,..., v_M\}$ and $M < m$, and corresponding contribution rates are $p_1,...., p_M$ and $\sum_{i=1}^{M} p_i > 85\%$, where $p_i = \dfrac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}$, and $T_{m\times M}$ will be used in testing process.

d) The output matrix

Suppose the output matrix is $New\_Y_{n\times M}$, and it can be obtained through (5).

$$New\_Y_{n\times M} = Y_{n\times m} \cdot T_{m\times M} = \left\{ \begin{matrix} y_{11}\ y_{12} \cdots y_{1M} \\ \cdots \\ y_{m1}\ y_{m2} \cdots y_{mM} \end{matrix} \right\}_{m\times M} \quad (5)$$

5) Health indicator

The prognostics approach in this paper is similar to [13], while the method in [13] only used instance learning to get the RUL, while in this paper, the HI and instance learning are both used. HI is widely used [7-9], and HI can be defined as a set of features extracted from monitored component which represent the component's degradation evolution as a function of time [5]. Here, the HI is actually a 1-D time series, and the method to get it is as follows.

Firstly, with the help of corresponding contribution rates $\{p_1,...., p_M\}$, we get each vector's weight of New_Y_{n*M} as follows:

$$w_i = \frac{p_i}{\sum_{j=1}^{M} p_j} \quad (6)$$

The weight vector is $W = \{w_1,..., w_M \mid w_1 + ... + w_M = 1\}$. And the WED vector $D_{m\times 1} = \{d_1, d_2,..., d_n\}$ can be obtained as:

$$d_j = \sqrt{\sum_{i=1}^{M} w_i {y_{ij}}^2} \ , \ j = 1,2,..., n \quad (7)$$

At last, the HI is usually a continuous-value quantity, so we need to regress the WED to get the HI curve. The methods of regression vary with the data, such as such as Kernel Regression [10], Locally Weighted Scatter plot Smoothing (LOWESS) [11], Relevance Vector Machines [12] and so on.

6) Failure Threshold

As the end value of each train unit is independence and does not impact relatively, and obey the same distribution. We suppose these end values obey normal distribution, the mean value can be selected as the threshold, and the confidence interval can be set as [μ-2σ, μ-2σ], where μ and σ are the mean and variance of the end values. What's more, the Jarque-Bera test method can be used to judge if matching a normal distribution. While the end values do not obey normal distribution, we can use the median value or mean value as the FT.

B. Testing process

1) Test data processing

The test data processing is similar to the train process, and it will not be repeated here. But the following points should be concerned. Firstly, in the normalization, the values of mean and standard deviation are the same to train data, and the same to transition matrix in dimension reduction and threshold value in FT.

2) RUL calculate

As mentioned previously, the methods to predict RUL can be divided into model-based method and data-driven method. In this paper, the RUL predicted method, which is based on instance learning, belongs to the data-driven method.

Specifically, if the train units' HI curves are about the same, then all units share the same HI curve. Only if this HI curve and FT are determined, the RUL can be determined. But in practice, the HI curves vary according to the operational conditions and other factors. Here, we give two approaches, trajectory similarity approach and FT matching approach, to calculate RUL and both of them are based on instance learning.

Trajectory similarity approach is detailed in [13]. In short, take each test unit to match all of the train HI curves and return the RULs, and the final prediction of RUL can be generate from these calculate RULs.

As to the FT matching approach, the threshold value is obtained from the train process, and the HI curve can be regressed by the test points. Suppose the value of FT is y_{FT} , and the HI curve can be expressed as $f_{HI}(t)$, the x value of last test point is x_{end} , then the RUL can be calculated from the (8).

$$RUL_{calculate} = Round(f_{HI}^{-1}(y_{FT}) - x_{end}) \quad (8)$$

Where, the function of 'Round' means rounding to nearest integer.

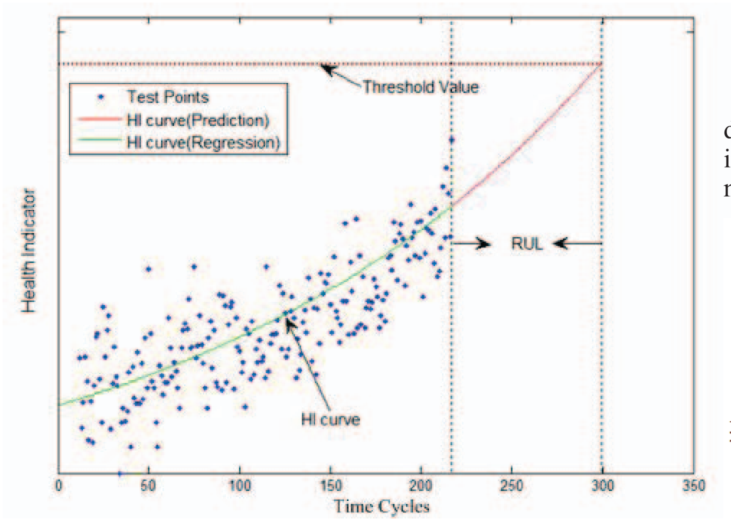

Figure 2. The FT matching approach and the RUL

III. CASE STUDY

A. Case description

The data used here supplied by NASA Ames is turbofan engine degradation simulation data set. The data set, divided into training and test subsets, is considered to be from a fleet of engines of the same type. In the training set, the trajectories are 100 and the fault grows in magnitude until system failure. In the test set, the trajectories are also 100 and the time series ends some time prior to system failure.

There are 26 columns totally in FD001. The condition and the fault mode is only one, so the data here we use is from Column 6 to Column 26. As to the remaining 21 columns, only 13 of them have the trend over time. So we select the 13 columns as Trending Data.

B. Training Process

1) Outliers Removal

As (1) ~ (4), we remove the outliers from the Trending Data. The value of p is 2 and k is 6. As the Unit 1 in the training set, the first column data is shown as Fig. 3 in the left part, y is the raw data and y1 is the mean data obtained from (1). In the right part, MAD is mean absolute deviation of y and y1.

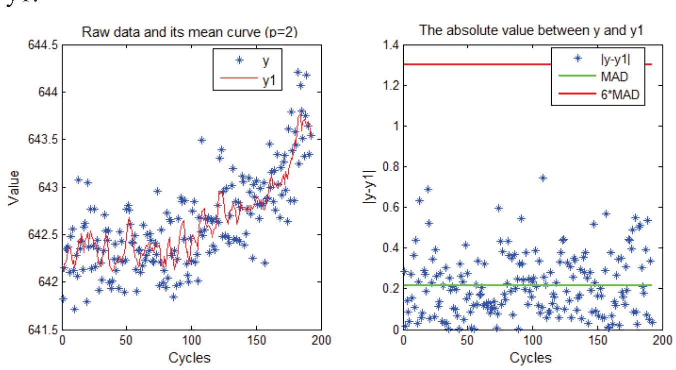

Figure 3. Outliers removal of Unit 1's first column

2) Normalize the train data

Z-score standardization is used to normalize the trending data, and the mean and variance of the 13 columns are shown in Fig. 4. The left is the mean and variance value before normalizing, and the right is after normalizing.

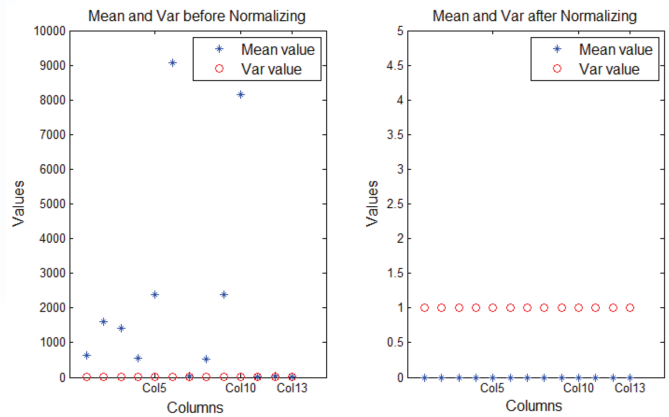

Figure 4. The mean and variance of Trending Data

After normalizing, all of the 13 columns have the same mean and variance, the mean value is zero and the variance is one. Fig. 5 shows the first column Unit1 of trending data, the left is before normalizing, and the right is after normalizing.

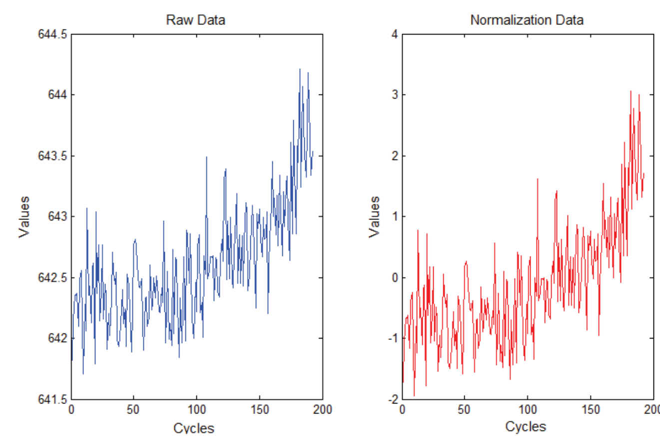

Figure 5. The data values before normalizing and after normalizing

3) Dimension reduction and HI obtained

After the PCA calculating, the dimension is reduced to two: the weight of first component is about 80%, and the second is 20%, and the Fig. 6 shows the weighted curve of Unit 1.

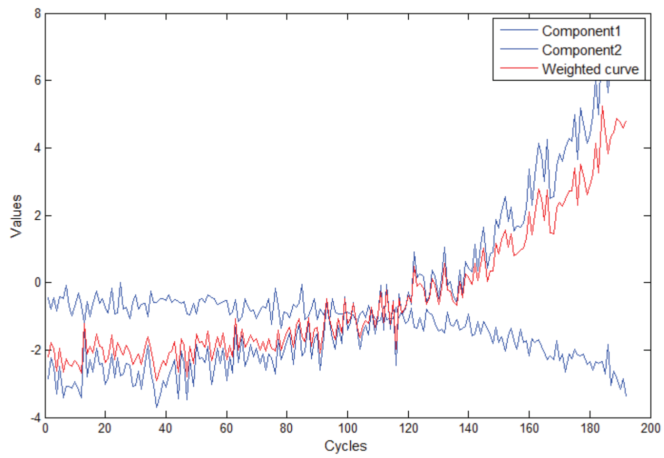

Figure 6. The weighted curve of Unit 1

The HI curve is obtained from the weighted curve by regression, the regression method here we used is exponential regression. The HI curves are shown as Fig. 7, the left part is the weighted curve (blue points) and the red is the HI curve. The right part is the HI curves of all train units.

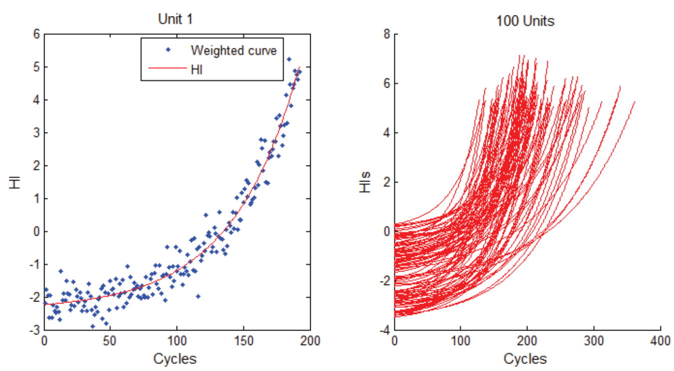

Figure 7. The HI curves of train units.

4) *FT calculation*

The end values of each HI curves are as shown in Fig. 8, and the end values do not obey normal distribution, then the median value or mean value can be used as the FT. As Fig. 8 shown, most of end values are between [4.87, 5.57], and the 90% confidence interval is [4.87, 6.64]. The mean value is 5.64 and the median value is 5.40. It is generally desirable to have early RUL estimates rather than late RULs, since the main aspect is to avoid failures [15], so the median value is more suitable to be selected as the threshold value.

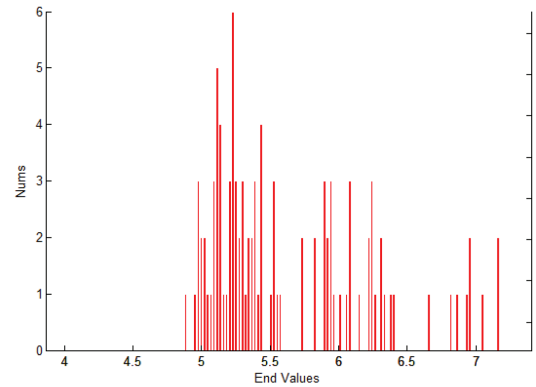

Figure 8. The end values of HI curves

C. *Testing process*

To calculate the RUL, two methods, trajectory similarity approach and FT matching approach, are both available. As trajectory similarity approach had been used in [14], so here we use FT matching approach to calculate the RUL for the test data.

To the FT matching approach, the more data, the more accurate. As the length of each test unit is different from others, and some of them are so short that it's hard to determine the parameters of HI functions. To solve this problem, we make use of the results of the train data HI curves which might restrict the parameters varying range.

The answer of the calculate RUL is shown in Fig. 9. In (a), the red line is the calculate RUL, and the blue is actual RUL, while in (b), the errors (actual values – calculate values) are shown and the values of red lines are 50 and -50. What's more, the distribution of the RUL errors is shown in Fig. 10.

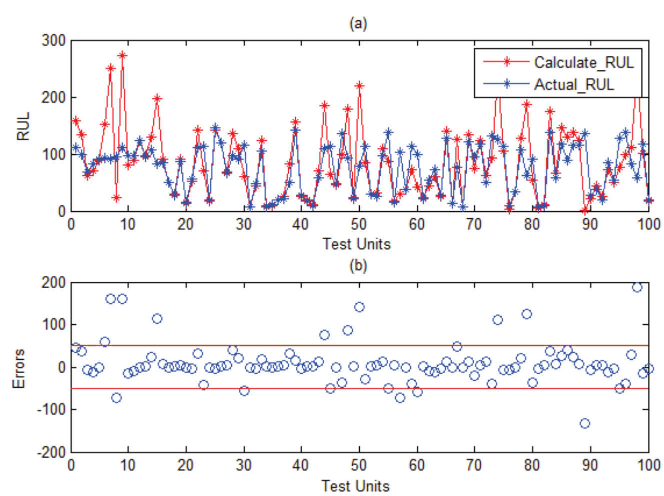

Figure 9. The calculations of the case study

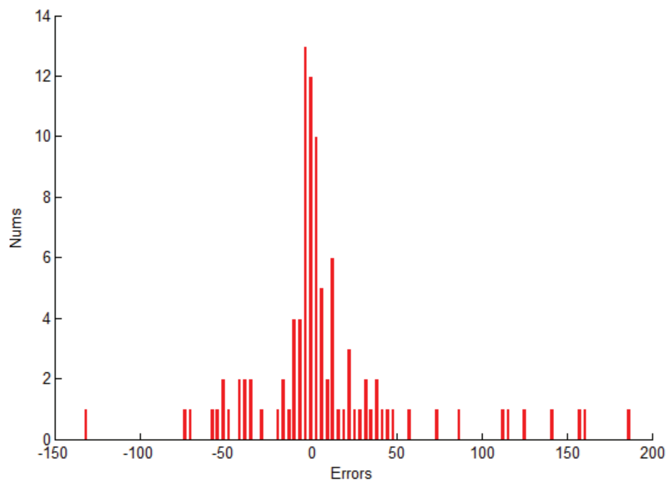

Figure 10. The distribution of errors

IV. Conclusion

In this paper, a data-driven approach for estimating RUL based on principle component and instance learning is introduced. Firstly, PCA is used to reduce dimension. Then, the HI curves can be obtained from PCs by using WED, and the FT can also be calculated. At last, the method based on instance learning is employed to estimate the RUL of the machine under operation. The method here for estimating RUL is especially suitable for situations in which abundant RtF data are available. However, the deficiencies could not be ignored. Firstly, the method to get the FT is lack of universal. Secondly, only the exponential distribution is discussed in FT matching approach. So, in the future, the work will be focus on solving the deficiencies, besides, more cases will be studied to show the effectiveness and generality of the proposed approach.

Reference

[1] A. Ray and S. Tangirala, "Stochastic modeling of fatigue crack dynamics for on-line failure prognostics," IEEE Transactions on Control Systems Technology, vol 4(4), pp. 443–451, July 1996.

[2] S. Marble and B. P. Morton, "Predicting the remaining life of propulsion system bearings," In 2006 IEEE Aerospace Conference, vol. 1-9, pp 4091–4098, Mar 2006.

[3] G. Box, G. M. Jenkins, and G. Reinsel, Time Series Analysis: Forecasting & Control, 3rd ed., Prentice Hall, February 1994.

[4] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE transactions on systems, man and cybernetics, vol. 15(1), pp. 116–132, 1985.

[5] A.Mosallam, K. Medjaher, N. Zerhouni, "Component based Data-driven Prognostics for Complex Systems: Methodology and Applications," The First International Conference on Reliability Systems Engineering. 2015, in press.

[6] I. B. Mohamad and D. Usman, "Standardization and Its Effects on K-means Clustering Algorithm," Research Journal of Applied Sciences, Engineering and Technology, vol. 6(17), pp. 3299-3303, 2013.

[7] R. Huang, L. Xi, X. Li, C. R. Liu, H. Qiu, and J. Lee, "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods," Mechanical Systems and Signal Processing, vol. 21(1), pp. 193–207, Jan 2007.

[8] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, and J. Lee, "Similarity based method for manufacturing process performance prediction and diagnosis," Computers in Industry, vol. 58(6), pp. 558–566, Aug 2007.

[9] Cheng and M. Pecht, "Multivariate state estimation technique for remaining useful life prediction of electronic products," In AAAI Fall Symposium on Artificial Intelligence for Prognostics, pp. 26–32, Nov 2007.

[10] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2nd ed., Prentice Hall, 2003.

[11] P. Craven and G. Wahba, "Smoothing noisy data with spline functions – estimating the correct degree of smoothing by the method of generalized cross-validation," Numerische Mathematik, vol. 31(4), pp. 377–403, 1979.

[12] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," Journal of Machine Learning Research, vol. 1(3), pp. 211–244, 2001.

[13] T.Wang, Trajectory Similarity Based Prediction for RUL Estimation, PhD thesis, University of Cincinnati, 2010.

[14] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," In Procee dings of the 2008 International Conference on Prognostics and Health Management, pp. 1–6, Denver, Oct 2008.

[15] K. Javed, R. Gourivean, and N. Zerhouni, "A new multivariate approuch for prognostics based on extreme learning machine and fuzzy clustering," IEEE Transactions on Cybernetics, vol. 45, pp. 2626-2639, Dec 2015.

V. Appendix

A. List of abbreviations

RUL-Remaining Useful Life

PHM- Prognostics and Health Management

RtF-run-to-failure

PC-principal component

WED -Weighted Euclid distance

HI-health indicators

FT-failure threshold

MAD-mean absolute deviation

B. The analysis of covariance between Min-Max scaling and Z-score standardization

Suppose $X=\{x_1,x_2,...,x_n\}$, $Y=\{y_1,y_2,...,y_n\}$ Then the mean, variance and covariance can be obtained as following equations.

$$E(X)=u_1\,, D(X)=\sigma_1^2\,; E(Y)=u_2\,, D(Y)=\sigma_2^2$$

$$Cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-u_1)(y_i-u_2)$$

As to Min-Max scaling,

$$x_i^{"}=\frac{x_i-\min(X)}{\max(X)-\min(X)}\,, y_i^{"}=\frac{y_i-\min(Y)}{\max(Y)-\min(Y)}$$

$$Cov(X'',Y'') = \frac{1}{n}\sum_{i=1}^{n}(x_i^{''} - u_1^{''})(y_i^{''} - u_2^{''})$$
$$= \frac{1}{ab}Cov(X,Y)$$

Where, a = max(X) – min(X), b = max(Y) – min(Y).

The covariance after normalizing is 1/ab of the original. Thus, the Min-Max scaling cannot eliminate the influence to the variance and covariance.

As to Z-score standardization, the covariance after normalizing can be calculated as follows.

$$x_i^{'} = \frac{x_i - \mu_1}{\sigma_1}, \quad y_i^{'} = \frac{y_i - \mu_2}{\sigma_2}$$

$$Cov(X',Y') = \frac{1}{n}\sum_{i=1}^{n}(x_i^{'} - 0)(y_i^{'} - 0)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \mu_1}{\sigma_1}\frac{y_i - \mu_2}{\sigma_2}$$
$$= \frac{1}{\sigma_1\sigma_2}Cov(X,Y)$$

As each principle component obeys standard normal distribution, Z-score standardization eliminates the influence to the variance and covariance.