Sensor Selection with Grey Correlation Analysis for Remaining Useful Life Evaluation

Peng Yu, Xu Yong, Liu Datong, Peng Xiyuan

Automatic Test and Control Institute, Harbin Institute of Technology, Harbin, 150001, China

pengyu@hit.edu.cn
xuyong.hit@gmail.com

ABSTRACT

Sensor selection in data modeling is an important research topic for prognostics. The performance of prediction model may vary considerably under different variable subset. Hence it is of great important to devise a systematic sensor selection method that offers guidance on choosing the most representative sensors for prognostics. This paper proposes a sensor selection method based on the improved grey correlation analysis. From empirical observation, all the continuous-value sensors with a consistent monotonic trend are firstly selected for data fusion, and a linear regression model is used to convert the multi-dimensional sensor readings into one-dimensional health factor (HF). The correlation between HF and each of the selected sensors is evaluated by calculating the grey correlation degree defined on two time series. The optimal sensor subset with a relatively large correlation degree is selected to execute the final fusion. The effectiveness of the proposed method was verified experimentally on the turbofan engine simulation data supplied by NASA Ames, using instance-based learning methodology, and the experimental results showed that RUL prediction with fewer sensor inputs can obtain a more accurate prognostics performance than using all sensors initially considered relevant.

1. INTRODUCTION

Prognostics and health management (PHM) of complex engineered systems has gained increasing attention from the research community worldwide. Prognostic builds the foundation of PHM, and its outcome directly affects the other PHM components such as operations planning, timely maintenance, logistics, etc. Hence prognostics can play an important part in reducing cost, increasing safety, and accomplishing critical missions (Heng et al., 2008). Generally, prognostics can be divided into the detection of failure precursors and the prediction of remaining useful life (RUL). Comparing to the detection of failure precursors, RUL prediction is mostly irrelevant to the application. The methods for RUL prediction are almost the same to all prognostics applications.

The ultimate aim of most prognostics systems is accurate estimation the RUL of individual systems, and the prediction accuracy relies not only on the prediction models used, but also on the types and number of sensors selected (Cheng et al., 2010). As the performance of system degrades, the monitored parameters tend to change accordingly. These raw multi-dimensional sensor data or features extracted from them may be used to track the degradation behavior of system. Typically, these degradation data can be used as the inputs of data-driven prognostics model to make RUL estimation. Degradation data may consist of sensor readings, such as temperature and pressure, or inferred features, such as model residuals or physics-based model predictions. Commonly, it is beneficial to fuse sensor readings and inferred features into a single health factor, which is considered as a more robust input to the prognostics model (Coble, 2010). However, inclusion of irrelevant or redundant variables during the fusion may lead to over-fitting or less sensitivity of prognostics model, which is adverse to the prediction performance. Hence variable selection is critical to make an accurate RUL estimation. Typically, sensor selection is left to expert knowledge, empirical observation of available data, and intimate knowledge of degradation mechanisms. These methods are time-consuming, and scale with the number of available sensors and possible fault modes (Zhang, 2005).

For many real-world systems, it is almost impossible to fully understand the degradation behavior of the systems and employ the first-principle models for prognostics. Since Instance-based learning (IBL) approach develops prognostics model based on a mass of historical instances, it becomes a preferable choice (Xue et al., 2008). As the rapid development of communication and sensor technology, abundant data collection from complex systems, such as aircraft engines, satellite power system, etc, becomes possible. These massive life-cycle condition data collected from various instances of the equipment further promotes the application of IBL prognostics approach.

The current work focuses on improving the RUL prediction accuracy of complex systems via sensor selection. In real applications, complex system is operated under dynamic operating conditions, and k-means clustering algorithm is employed to cluster the operational conditions into a finite number of operating regimes. According to the empirical observation, those sensors with a consistent monotonic trend under different operating regimes are selected, and a linear regression model is employed to convert the selected multivariate sensor readings from individual regimes into a one-dimensional HF, then all the HFs are merged to form a complete HF time series with the original time stamps. The correlation between HF and each of the selected sensors is evaluated by calculating the grey correlation degree defined on two time series. The optimal sensor subset with a relatively large correlation degree is selected to execute the final fusion. Finally, the obtained HF time series is integrated into an IBL prognostics architecture consisting of model recognition, similarity evaluation and RUL prediction. Moreover, the performance of the sensor selection strategy is verified experimentally on the turbofan engine simulation data supplied by NASA Ames, using the integrated IBL prognostics architecture.

The paper is organized into the following sections. In section 2, the sensor selection scheme based on improved grey correlation analysis is introduced. In section 3, the IBL prognostics architecture is summarized. The turbofan engine application is elaborated in section 4. The experiment results and discussions are presented in section 5. Conclusion is drawn in section 6.

2. SENSOR SELECTION SCHEME

Sensor selection is mostly relevant to the application, and it aims at reducing the unnecessary redundancy while maximizing the relevance in the sensor subset (Wang et al, 2008). According to the characteristics of collected data, the sensor selection scheme probably consists of operation condition division, empirical observation, data fusion and grey correlation analysis.

2.1. Operation Condition Division

In real applications, the dynamic operation conditions have a great impact on the sensor readings or inferred features from the system and complicate the system degradation behaviors over time. The sensor time series may show little trend. However, if the operation conditions are clustered into several operating regimes by the use of certain clustering algorithm, the sensor data collected from different regimes may exhibit a prominent trend.

The vector $\mathbf{c}_i$ represents the operation conditions of the system at time t_i. Suppose the operational conditions $\mathbf{c}$ can be concentrated into a limited number of operating regimes $\mathbf{O}=\{O_1,O_2,...,O_P\}$ using k-means clustering algorithm f. The output of f is defined as:

$$\mathbf{S}=f(\mathbf{c})=(S_1,S_2,...,S_P) \quad (1)$$

where S_p is membership score when $\mathbf{c} \in O_p$.

In case of discrete operation conditions, the output of f can be simplified as:

$$f(\mathbf{c})=\arg\max_{k=1,...,P} S_k \quad (2)$$

2.2. Empirical Observation

The objective of empirical observation is to eliminate the sensors which are obviously not suitable for prognostics modeling. The visual inspection procedures of sensor data under different operating regimes are as follows:

1. Some sensors with one or multiple discrete values are firstly discarded, from which it is difficult to track the degradation trend of the system.
2. Some other sensors have continuous values, but exhibit non-monotonic trend during the life time of the instances, should also be discarded.
3. All the remaining sensors with continuous values exhibit a monotonic trend, but some of them show inconsistent evolution trend among the different instances, which may represent various fault modes of the system. As it is hard to quantize or identify the fault modes without system relevant information, those sensors with inconsistent trend are eliminated.
4. Only those sensors with a consistent monotonic trend under different operating regimes are selected for data modeling or other processing.

2.3. Data Fusion

In this paper, data fusion refers to convert the selected multivariate sensor data from individual regimes into a single HF within a normalized range. Therefore, the HFs obtained from each regime can be merged to form a new one-dimension time series, as described in Fig.1.

For operating regime O_p, one local regression model is created:

$$z = h\ (\mathbf{x};\boldsymbol{\theta}^{(p)}) \quad (3)$$

where $\mathbf{x}$ represents the multivariate sensor data, and $\boldsymbol{\theta}^{(p)}$ denotes the local model parameters.

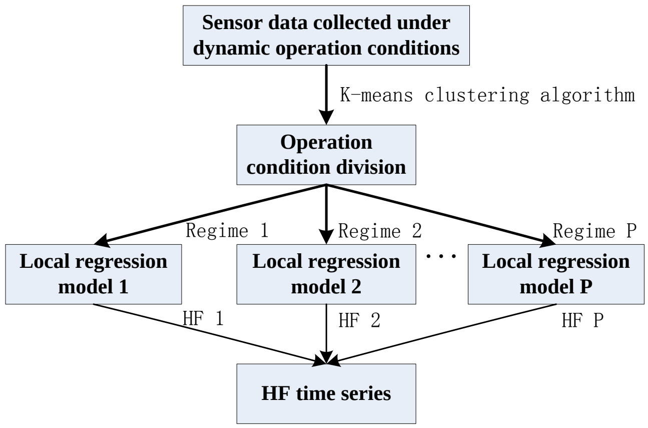


Figure 1. Data processing for RUL modeling through multi-regime health assessment

To make the HF comparable under different operating regime, the obtained HF should be normalized to a range, usually between 0 and 1, hence, the learning procedures of a local regression model can be followed:

- For those samples $(\mathbf{c}_i, \mathbf{x}_i)$ collected at early stage, that is $t_i < T_1$, assign the matching output $z_i = 1$;
- For those samples $(\mathbf{c}_i, \mathbf{x}_i)$ collected at middle stage, that is $T_1 < t_i < T_2$, will not participate model training;
- For those samples $(\mathbf{c}_i, \mathbf{x}_i)$ collected at late stage, that is $t_i > T_2$, assign the matching output $z_i = 0$.

The number and parameters of local models vary with the assigned thresholds T_1 and T_2, which can greatly influence the model performance. Commonly, the parameters are chosen by the rule of thumb, e.g. $T_1 = t_E*10\%$ and $T_2 = t_E*90\%$, where t_E represents the whole lifespan of the instance. Above all, once the sensor data of training instances $\mathbf{x}_i$ are provided, they will be divided into finite groups in accordance with different operating regimes and applied to learn different local regression models.

2.4. Grey Correlation Analysis

As mentioned, the sensors are preliminary selected by empirical observation, and the sensor fusion method is provided. One major problem lies in whether the sensor subset can be further optimized. Due to large noise and low sensitivity, some sensors exhibit an unclear trend compared with the others. Including them in the data fusion may lower the prediction accuracy. Hence certain analysis method should be adopted to further select the sensors.

Grey correlation analysis is a principle theory of grey system theory, which can be applied in grey system analysis and random variables processing (Zhang & Zhang, 2007). The correlation between factors is represented by the similarity level of geometry which is called grey correlation degree, and the correlation degree between reference sequences and comparison sequences can be quantitatively estimated. Grey correlation degree describes the relative change between different factors in the process of system evolution, and the larger the correlation degree is, the higher the similarity level is. Thus, the correlation degree can represent the impact of different sensors on the system degradation behaviors. By the calculation of improved correlation degree between HF time series and sensor time series, the sensors with a relatively large correlation degree are selected since they have a larger impact on the HF. The calculation steps of grey correlation degree between HF and sensors are as follows:

1. The HF time series which can represent the system degradation behaviors is set as the reference sequences $\mathbf{Z} = \{z(k) | k = 1, 2, ..., n\}$, and the sensor time series which can affect the system degradation behaviors are set as the comparison sequences $\mathbf{X}_i = \{X_i(k) | k = 1, 2, ..., n\}$, $i = 1, 2, ..., m$.

2. Due to the various units of measurements, the dimensions of sensor data are different, which may lead to a wrong correlation analysis result. Thus, the data should be converted into dimensionless form. There are many dimensionless processing methods, such as equalization, initialization, etc. In this paper, initialization method is applied, and the whole data in the original sequences are divided by the initial data, which is shown as follows:

$$x_i(k) = \frac{X_i(k)}{X_i(1)},\ k = 1, 2, ..., n;\ i = 1, 2, ..., m \tag{4}$$

3. The correlation is substantially the fitting degree of geometry between curves, thus the difference between curves is considered as the performance indicator of correlation. Set $\Delta_i(k) = |z(k) - x_i(k)|$, the grey correlation coefficient between $z(k)$ and $x_i(k)$ is:

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)} \tag{5}$$

   where $\rho \in (0, \infty)$, and its effect lies in enhancing the significance of difference between the correlation coefficients. The value of ρ is commonly set as 0.5.

4. The correlation coefficient represents the correlation between reference sequences and comparison sequences at various points, and there is a corresponding correlation coefficient at each point. However, the decentralized information is inconvenient for holistic comparison. To solve the problem, the correlation degree is proposed, which can be represented by the mean of the correlation coefficients:

$$r_i = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k), k = 1,2,...,n \qquad (6)$$

5. The above-mentioned correlation degree is calculated without considering the diversity of correlation coefficients at different point. Therefore, the stability of correlation coefficient sequences is proposed:

$$S(r_i) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\xi_i(k) - r_i)^2} \qquad (7)$$

On the basis of the stability, the computing model of grey correlation degree is improved:

$$r^*(i) = \frac{r_i}{1 + S(r_i)} \qquad (8)$$

The object of calculating grey correlation degree lies in comparing the impact of different sensors on system degradation behaviors. If $r_1 > r_2$, the comparison sequence $\mathbf{X}_1$ has a greater impact on $\mathbf{Z}$ than $\mathbf{X}_2$.

3. IBL Prognostics Architecture

From the above-mentioned sensor selection scheme, an optimal sensor subset is selected, and HF time series is formed by sensor data fusion, which is integrated into IBL prognostics architecture. In this architecture, a number of HF time series extracted from the historical monitoring data of the training instances with known failure times are used to form a library of degradation models. The similarity between a test instance and each of the models are evaluated, and each model can give an individual RUL estimation to the test instance. These RUL estimations can fuse into a final RUL prediction by the similarity-weighted sum. The architecture consists of model recognition, similarity evaluation and RUL prediction.

3.1. Model Recognition

The HF time series extracted from one training instance is available to establish a model depicting the whole performance degradation process of the instance, and the model library $\{M_i\}$ can be constructed based on the multiple HF time series extracted from training instances. M_i is commonly a deterministic model that can give a predicted output at a given time:

$$M_i : y = m_i(t) + \varepsilon, -T_i \le t \le 0 \qquad (9)$$

where ε is noisy term, T_i is the lifetime of the training instance used to establish the model.

The selection of model type is application dependent. For complex engineered systems, the main consideration is commonly focused on the long term degradation trend of the system, and the fluctuations in the degradation process can be recognized as interference or noise. Hence, a model with smoothing function of the time series can be adopted.

3.2. Similarity Evaluation

The definition of similarity between different instances has a great impact on the performance of IBL prognostics method. In this paper, grey correlation degree and Euclidean distant are adopted to respectively represent the similarity between the test instance $\mathbf{Z}_T = \{z(k) | k = 1,2,...,r\}$ and degradation model M_i. The calculation steps of grey correlation degree between different data sequences have already been introduced in section 2.4 and will not be reiterated here.

The Euclidean distant between $\mathbf{Z}_T$ and M_i is defined as:

$$D(\tau, \mathbf{Z}_T, M_i) = \sum\nolimits_{j=1}^{r}(z(j) - m_i(-\tau - r + j))^2 / \sigma_i^2 \qquad (10)$$

where $0 \le \tau \le T_i - r + 1$, τ represents the time span that the time series $\mathbf{Z}_T$ is moved away from cycle zero of model M_i, and σ_i^2 is the prediction variance provided by M_i. The smaller the distance is, the higher the similarity is.

Moreover, the similarity degree between $\mathbf{Z}_T$ and M_i is defined as:

$$S(\tau, \mathbf{Z}_T, M_i) = \exp\left(-D\left(\tau, \mathbf{Z}_T, M_i\right)\right) \qquad (11)$$

3.3. RUL Prediction

Once the definition of similarity is defined, each model M_i in the library can give an individual RUL estimation to the test instance:

$$RUL_i = T_i \text{ - } r \text{ - } \arg\min_{\tau} D(\tau, \mathbf{Z}_T, M_i) \qquad (12)$$

All RUL predictions and corresponding similarity degrees form a set $\left\{\left(RUL_i, S\left(\tau, \mathbf{Z}_T, M_i\right)\right) | i = 1,2,...,I\right\}$, where I represents the number of models in the library. The similarity-weighted method is applied to fuse all the RUL predictions in the set to get a final RUL estimation:

$$RUL_r = \frac{\sum_{i=1}^{I} S(\tau, \mathbf{Z}_T, M_i) \cdot RUL_i}{\sum_{i=1}^{I} S(\tau, \mathbf{Z}_T, M_i)} \qquad (13)$$

4. Case Study

In this section, the performance of the sensor selection scheme will be validated experimentally on the turbofan engine simulation data available from NASA Ames Prognostic Data Repository (Saxena & Goebel, 2008), using the integrated IBL prognostics architecture.

	Set 1	**Set 2**	**Set 3**	**Set 4**
Fault modes	1	1	2	2
Operation condition	1	6	1	6
Training units	100	260	100	248
Testing units	100	259	100	249

Table 1. Experiment settings of the data sets

4.1. Data Description

Four data sets with different simulation settings such as the number of operation conditions and fault modes are provided by NASA, and these data sets consist of multivariate time series from multiple instances of the turbofan engine. The collected data for each instance consists of a 24-dimensional time series (3 operation conditions and 21sensor readings for each flight cycle), and these data can represent the condition of engine throughout its flight history. The experiment settings of the data sets are described in Table 1.

Furthermore, each data set is divided into training and testing subsets. The instances in the training subset have complete run-to-failure data, which can be applied to develop prognostics model, while the instances in the testing subset have up-to-date data and corresponding failure time data, which can be applied to validate the performance of prognostics model. In this paper, data set 2 is selected to evaluate the effectiveness of the proposed sensor selection scheme. Nevertheless, only the training subset is applied for validity assessment, for the testing instances with incomplete run-to-failure data are not suitable for performance evaluation metrics based on successive RUL estimations throughout the whole life. Hence, the first 200 training instances will be applied for training, and 30 out of the 60 remaining training instances will be selected randomly for testing.

4.2. Performance Metrics

In the context of prognostics, the traditional accuracy-based or robustness-based performance metrics are inadequate to fairly assess the performance of prediction algorithms. Hence, four performance metrics proposed by Saxena et al. (2010) are adopted with minor modification. These metrics are on the basis of successive RUL prediction for each instance.

1) Prediction horizon

Prediction horizon (PH) is defined as the RUL estimation that firstly satisfies the α -bound criteria:

$$\text{PH}=t_E\text{-}t_\alpha \tag{14}$$

where t_E is the end-of-life time stamp, and t_α represents the time stamp of the RUL estimation that firstly satisfies the α -bound criteria. At each time stamp t_i , the corresponding RUL estimation is r_i . t_α is defined as:

$$t_\alpha = \min\{t_i | t_i \in [t_s, t_f],\ r_i^*\text{-}\alpha \cdot t_E \le r_i \le r_i^* + \alpha \cdot t_E\} \tag{15}$$

where t_s is the start time of the RUL estimation, and t_f denotes the end time of the RUL estimation.

2) Rate of acceptable predictions

Rate of acceptable predictions (AP) is defined as the rate of predictions that fall into an acceptable cone-shape area when $t_i \ge t_h$:

$$\begin{aligned} &\text{AP=Mean}(\{\delta_i | t_h \le t_i \le t_f\}) \\ &\delta_i = \begin{cases} 1, & \text{if } (1\text{-}\alpha) r_i^* \le r_i \le (1\text{-}\alpha) r_i^* \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{16}$$

where t_h is chose as PH calculated above. Obviously, AP is a stricter metric than the prediction errors.

3) Relative accuracy

Relative accuracy (RA) is defined as the mean absolute percentage errors for all $t_i \ge t_h$:

$$\text{RA=1-Mean}(\{\frac{|r_i\text{-}r_i^*|}{r_i^*} | t_h \le t_i \le t_f\}) \tag{17}$$

RA can give a quantitative metric of the prediction accuracy within the specified RUL, comparing with AR.

4) Convergence

Convergence (CG) evaluate how fast the prediction performance improves when more historical data is available:

$$\text{CG=1-}\left(\frac{\frac{1}{2}\sum_{i=s}^{f}(t_{i+1}{}^2\text{-}t_i{}^2)E_i}{\sum_{i=s}^{f}(t_{i+1}\text{-}t_i)E_i} \text{-}t_s \right) \cdot \frac{1}{t_f\text{-}t_s} \tag{18}$$

where performance metric $E_i = r_i\text{-}r_i^*$. The value of CG is between 0 and 1, CG>0.5 indicates convergence for the prediction.

5) Performance evaluation

To assess the prediction performance based on multiple series from K testing instances, the median of four performance metrics is used:

$$\begin{aligned} &\text{PH=Median}(\{^k\text{PH}\}_K) \\ &\text{AP=Median}(\{^k\text{AP}\}_K) \\ &\text{RA=Median}(\{^k\text{RA}\}_K) \\ &\text{CG=Median}(\{^k\text{CG}\}_K) \end{aligned} \tag{19}$$

Within the four metrics, PH represents the size of time interval while the others have a value between 0 and 1 (1 means perfect).

4.3. Training Stage

From visual inspection, the sensor data in the training instances exhibit no prominent trend, as shown in Fig.2. Hence, k-means clustering algorithm is employed to cluster the 3 operation conditions of all the 200 training instances, and the operation conditions are clustered in 6 discrete operating regimes, labeled by an ID from 1 to 6. By this way, the sensor data from each regime may exhibit a rising trend, as shown in Fig.3

On the basis of the empirical observation mentioned in section 2.2, the 9 sensors with a consistent monotonic degradation trend under all of the 6 operating regimes are selected for further processing, namely #2, #3, #4, #7, #11, #12, #15, # 20 and #21. For instance, the readings of sensor #2 from all the training instances under regime 1 are illustrated in Fig.4.

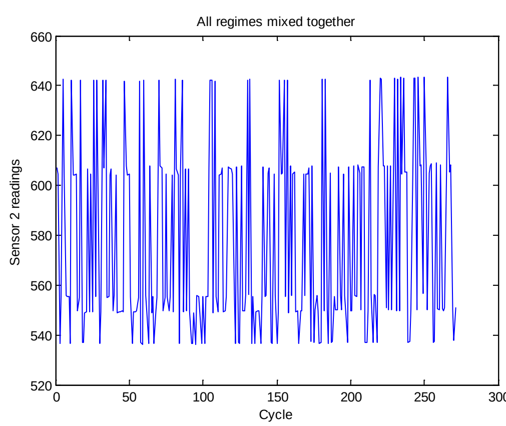


Figure 2. Raw data of sensor 2 from one training instance under all regimes

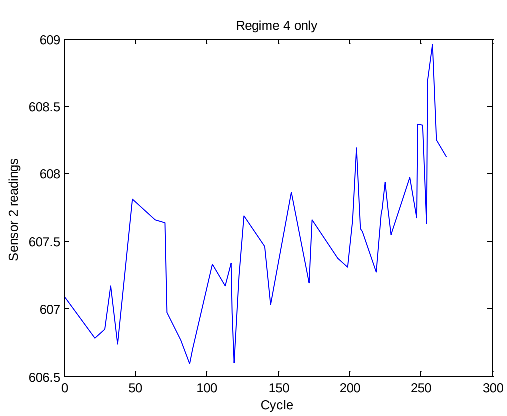


Figure 3. Raw data of sensor 2 from one training instance under regime 4 only

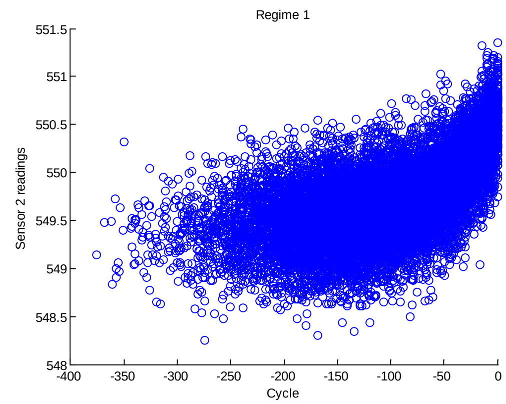


Figure 4. Raw data of sensor 2 from all the training instances under regime 1

Since the sensors have been preliminary selected, HF time series can be obtained through data fusion. It is remarkable that the selected sensors exhibit a consistent monotonic trend under different operating regimes, and a linear model can fit well for data with consistent trend. Hence, a linear regression model is adopted to convert the multi-dimensional sensor data into HF:

$$z = \alpha + \boldsymbol{\beta}^{\mathrm{T}} \cdot \mathbf{x} + \varepsilon = \alpha + \sum\nolimits_{i=1}^{9} \beta_i x_i + \varepsilon \qquad (20)$$

where $\mathbf{x}$ represents the selected 9-dimension sensor data, z is the health factor, and ε is the noise term.

The sample set $\boldsymbol{\Omega} = \{(\mathbf{x}, z)\}$ is used to learn the linear model:

$$\boldsymbol{\Omega} = \{\langle \mathbf{x}_i, 0\rangle \mid t_i > T_2\} \cup \{\langle \mathbf{x}_i, 1\rangle \mid t_i < T_1\} \qquad (21)$$

In this paper, the thresholds are set as T_1=-t_E*90% and T_2=-t_E*10%, where t_E represents the whole lifespan of the instance. Finally, 162 HF time series are extracted from all the 200 training instances, one of the time series is showed in Fig. 5.

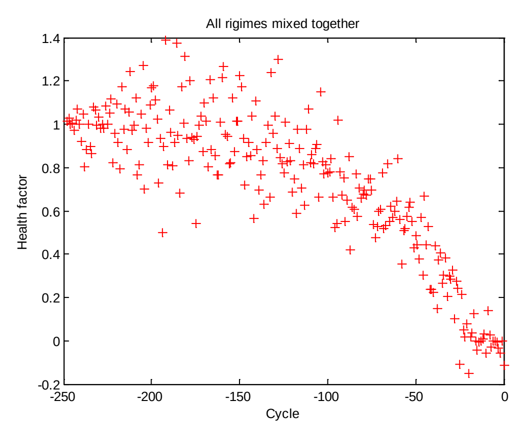


Figure 5. HF time series extracted from one training instance

Ranking	#203	#205	#206	#209
1	1(1)	1(1)	1(1)	1(1)
2	0.73636(6)	0.74473(6)	0.76490(8)	0.78437(8)
3	0.72580(8)	0.74145(8)	0.76392(6)	0.77928(6)
4	0.72313(3)	0.73965(3)	0.75887(3)	0.77731(2)
5	0.72121(2)	0.73890(2)	0.75671(2)	0.77700(3)
6	0.71695(4)	0.73636(4)	0.75041(4)	0.77345(4)
7	0.55575(10)	0.56147(10)	0.55503(9)	0.66692(9)
8	0.55283(9)	0.55972(9)	0.55435(10)	0.66538(10)
9	0.54224(5)	0.55115(5)	0.54571(5)	0.65627(5)
10	0.54137(7)	0.55020(7)	0.54537(7)	0.65556(7)

Table 2. Computing results of correlation degrees in the selected 4 training instances

The HF time series and corresponding 9-dimension sensor time series are labeled by an ID from 1 to 10. The grey correlation degrees between 162 HF time series and their corresponding multivariate sensor time series are calculated, and the results are sorted in descending order. Among them, the computing results of correlation degrees from 4 training instances are shown in Table 2.

From Table 2, the maximal correlation degree equals to 1, namely HF time series has a perfect similarity with itself. The correlation degrees ranked from 2 to 6 are rather close to each other, and so are the correlation degrees ranked from 7 to 10, but the correlation degrees between these two subsets have an obvious difference in value. It means that the first 5 sensor time series have a greater impact on the system degradation behaviors than the latter 4. By now, the first 5 sensors, namely #2, #3, #4, #11 and #15, might be selected to further optimize the sensor selection. To verify the generality of grey correlation analysis, the statistical information related with correlation degrees are given in Table 3.

From Table 3, the correlation degrees of two sensors #15 and #11, are ranked from 2 to 3 in most cases, the between HF and each of the sensors in all training instances correlation degrees of the sensors #2 and #3, are always ranked from 4 to 5, and the correlation degree of sensor #4 is ranked as 6 in most cases. Meanwhile, the correlation degrees of the first 5 sensors are rather close in value. Moreover, the correlation degrees of the remaining 4 sensors have never featured in the top 6 rankings, and the sensors ranked in the top 6 have an obviously higher coefficient degree than the sensor ranked in 7. Hence, the similarity between HF time series and the sensor time series can be well represented by the improved grey correlation degree, and the selected sensor subset includes 5 sensors, #2, #3, #4, #11 and #15. Furthermore, whether the selected optimal sensor subset can be further reduced remains to be discussed.

Ranking	#2	#3	#4	#7	#11	#12	#15	#20	#21
2	0	0	0	0	28	0	134	0	0
3	9	7	0	0	128	0	18	0	0
4	89	66	0	0	6	0	1	0	0
5	64	89	3	0	0	0	6	0	0
6	0	0	159	0	0	0	3	0	0
7	0	0	0	0	0	0	0	94	68
8	0	0	0	0	0	0	0	68	94
9	0	0	0	87	0	75	0	0	0
10	0	0	0	75	0	87	0	0	0

Table 3. Statistical information of correlation degrees

As seen in Fig.5, HF time series demonstrate an exponential degradation trend. Thus, the exponential regression models are adopted to describe the relationship between the HF z and operating time t:

$$z=a\cdot \exp(b\cdot t+c)+d+\sigma \tag{22}$$

where a, b, c, and d are the model parameters to be learned from HF time series, and σ is the noise term. The equation $z=0$ indicates the failure of the instance, which is equal to the constraint $a\cdot \exp(b\cdot t_E+c)+d=0$. After solving the parameter d, the following model form can be achieved:

$$z=a\cdot(\exp(b\cdot t+c)-\exp(b\cdot t_E+c))+\sigma \tag{23}$$

The model library $\{M_i\}$ can be constructed based on the 162 HF time series extracted from all the 200 training instances.

4.4. Testing Stage

The sensor data in the testing instances should be converted into HF time series. For each testing instance, the selected sensor data will be clustered by operating regimes, and transformed by the linear regression models obtained during the training stage, and fused to obtain a HF time series.

In this paper, the similarity between the test instance and each of the degradation models are respectively evaluated by grey correlation degree and Euclidean distance, using Eq. (8) and (11). The final point estimation of RUL is obtained using Eq. (13).

5. Performance Evaluation

For each testing instance, multiple RUL estimations will be made at different time stamps along the life of the instance, and each RUL prediction is made based on the up-to-date data till the corresponding time. In this application, the start time of prediction t_s is set to 50; the end time of prediction t_f is set to t_E-10; the time interval of prediction is set to 5.

1) Impact of the sensor selection scheme on prediction performance

By empirical observation, the sensors with a consistent monotonic degradation trend under all of the 6 operating regimes are selected, namely #2, #3, #4, #7, #11, #12, #15, # 20 and #21, which is represented by sensor subset 1. Furthermore, based on the grey correlation analysis, the optimal sensor subset 2 with a relatively large correlation degree is selected, and the subset includes 5 sensors, labeled by #2, #3, #4, #11 and #15. In this experiment, Euclidean distance is adopted to evaluate the similarity, and the thresholds of linear regression models are set as T_1=-t_E*90% and T_2=-t_E*10% . 30 testing instances are selected randomly to validate the impact of different sensor subsets on prediction performance, and the comparison results are described in Table 4.

As seen in Table 4, integrate with the IBL prognostics algorithm, the proposed sensor selection scheme improves the RUL prediction performance significantly. AP, RA and CG are improved by 25.9%, 4.3% and 1.6% while PH is nearly 10 cycles larger than before. Hence, the sensor selection scheme based on improved grey correlation analysis can effectively improve the RUL prediction performance. The sensor subset 2 will be adopted for data fusion in the subsequent experiments

2) Impact of different threshold settings for linear regression models on prediction performance

Not all the multivariate sensor data series can be converted into corresponding HF time series through linear regression. The number and parameters of linear regression models will vary with the different threshold settings. In this experiment, Euclidean distance is adopted to evaluate the similarity, and the thresholds of linear regression models are set to T_1=-t_E*90% and T_2=-t_E*10% , T_1=-t_E*95% and T_2=-t_E*5% , and T_1=-240 and T_2=-20 respectively. 30 testing instances selected here is identical to that of the previous experiment, which are used to validate the impact of different threshold settings on prediction performance, and the comparison results are described in Table 5.

As seen in Table 5, when T_1=-t_E*90% and T_2=-t_E*10% , AP, RA and CG are much greater than those in the other two threshold settings, except that PH is relatively small. Moreover, it seems as if the setting of threshold parameters using a certain percentage of the total life of the instance can lead to a better prediction effect than the threshold parameters with fixed value. If hard threshold is applied, All the training instances whose lifetime are smaller than T_1, cannot be used to train the linear regression models, resulting in the decreasing of the model types in the model library.

Performance Metric	Subset 1	Subset 2
PH	134	144.5
AP	0.49515	0.62337
RA	0.78859	0.82270
CG	0.67914	0.68985

Table 4. Prediction performance of IBL prognostics algorithm under different sensor subsets

Performance Metric	T_1=-t_E*90% T_2=-t_E*10%	T_1=-t_E*95% T_2=-t_E*5%	T_1=-240 T_2=-20
PH	144.5	145	160.5
AP	0.62337	0.31743	0.29872
RA	0.82270	0.60669	0.58335
CG	0.68985	0.56431	0.53223

Table 5. Prediction performance of IBL prognostics algorithm under various thresholds

3) Impact of different similarity measurements on prediction performance

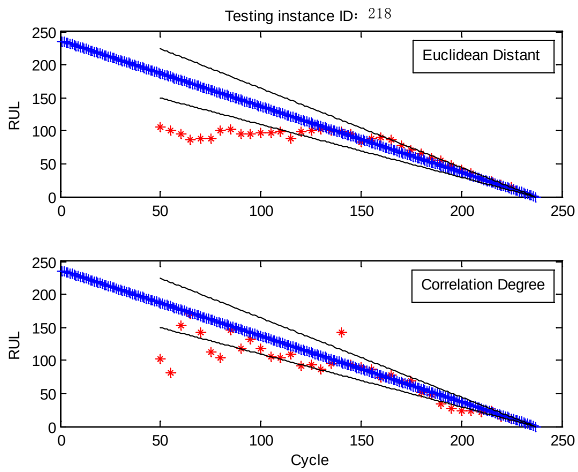
The definition of similarity between different instances has a great impact on the performance of IBL prognostics method. The purpose of this experiment is to figure out either Euclidean distance or grey correlation degree is a preferable similarity measurement. In the experiment, 30 testing instances which have been selected in the previous experiment are used to validate the impact of different similarity measures on prediction performance, and the comparison results are described in Table 6. RUL predictions of certain testing instances using different similarity measures are shown in Fig.6.

As seen in Table 6, combined in the IBL prognostics method, Euclidean distance is a preferable similarity measurement. AP, RA and CG are respectively improved by 19.3%, 8.8% and 3.2%, but PH is relatively smaller.

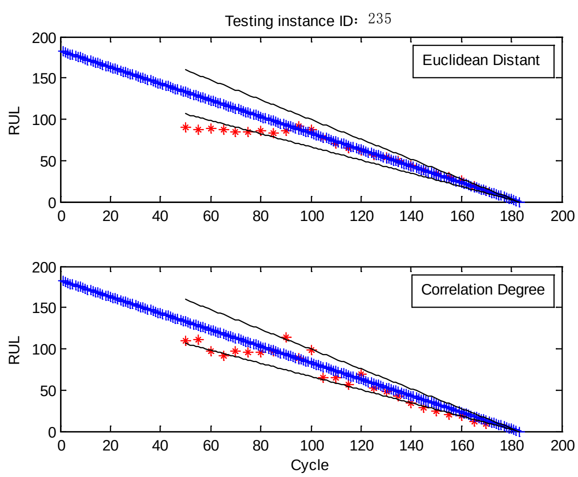
Above all, when sensor selection is executed based on grey correlation analysis, the threshold parameters are set to T_1=-t_E*90% and T_2=-t_E*10% , and Euclidean distance is chosen as the similarity measurement, the IBL algorithm can achieve an optimal RUL prediction performance.

Performance Metric	Euclidean Distance	Grey Correlation Degree
PH	144.5	161.5
AP	0.62337	0.52272
RA	0.82270	0.75639
CG	0.68985	0.63457

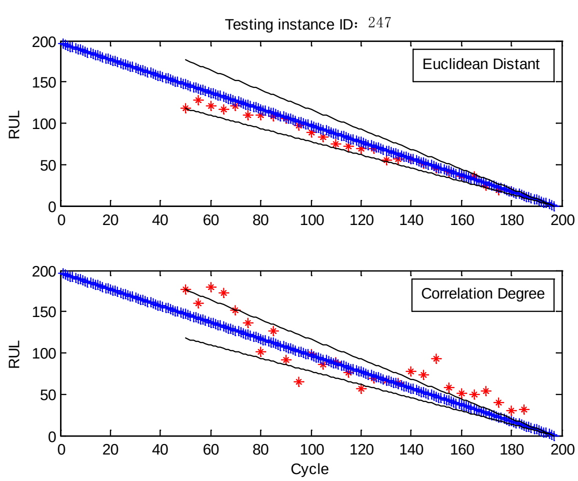
Table 6. Prediction performance of IBL prognostics algorithm under different similarity measurements



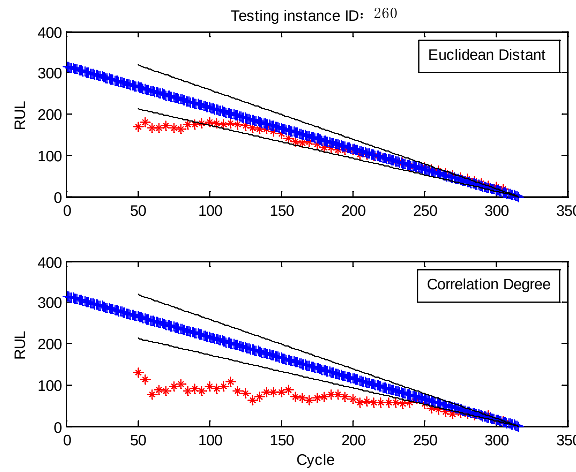
(a) Instance ID: 218



(b) Instance ID: 235



(c) Instance ID: 247



(d) Instance ID: 260

Figure 6. RUL predictions for selected instances using different similarity measures

From Fig.6, the selected instances show a desirable performance, where the RUL predictions converge to the true RUL as time increases. In the final stage, the RUL predictions are almost equal to the corresponding true RULs, indicating that the IBL algorithm has an excellent convergence and prediction effect in this application.

6. CONCLUSION

A grey correlation analysis method to selecting the most representative sensors for the RUL prediction of complex engineered systems is developed. The performance of sensor selection scheme integrated with IBL prognostics algorithm was evaluated using four performance metrics designed in the context of PHM. The addition of other sensors not selected by grey correlation analysis to the input sensor subset has led to a decreasing in the prediction performance, confirming the effectiveness of the sensor selection scheme. The scheme presented is expected to gain a similar performance for other complex engineered systems.

REFERENCES

Heng, A., Zhang, S., Tan, C. T., & Mathew, J. (2008). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, vol 23, pp. 724-739. doi:10.1016/j.ymssp.2008.06.009

Cheng, S., Azarian, M. H., & Pecht, M. G. (2010). Sensor systems for prognostics and health management. *Sensors*, vol 10, pp 5774-5797, doi:10.3390/s100605774

Coble, J. B., (2010). *Merging Data Sources to Predict Remaining Useful Life – An Automated Methods to Identify Prognostic Parameters*, Doctoral dissertation.

University of Tennessee, Knoxville, USA. http://trace.tennessee.edu/utk_graddis/683

Zhang, G. F., (2005). *Optimal Sensor Localization/Selection in A Diagnostic/Prognostic Architecture*, Doctoral dissertation. Georgia Institute of Technology, Atlanta, USA.

Xue, F., Bonissone, P., Varma, A., Yan, W. Z., & Goebel, K. (2008). An Instance-based method for remaining useful life estimation for aircraft engines. *Journal of Failure Analysis and Prevention*, vol 8, pp. 199-206. doi:10.1007/s11668-008-9118-9

Wang, T. Y., Yu, J. B., Siegel, D., & Lee, J. (2008). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. *International Conference on Prognostics and Health Management*. October 6-9, Denver, CO. doi:10.1109/PHM.2008.4711421

Zhang, Y. J., & Zhang, X. (2007). Grey correlation analysis between strength of slag cement and particle fractions of slag powder. *Cement and Concrete Composites*, vol 29, pp. 498-504. doi:10.1016/j.cemconcomp.2007.02.004

Saxena, A., & Goebel, K. (2008). C-MAPSS data set. *NASA Ames Prognostics Data Repository*.http://ti.arc.nasa.gov/project/prognostic-data-repository

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, vol 1.

BIOGRAPHIES

Peng Yu was born in 1973. He received the Ph.D. in instrument science and technology from Harbin Institute of Technology. Since 2007, he has been a professor in the Department of Instrument science and technology, Harbin Institute of Technology. His research interests include pattern recognition, data mining and wireless sensor networks.

Xu Yong was born in 1985. He is currently working toward the Ph.D. degree from Harbin Institute of Technology. His research interests include wireless sensor networks and prognostics and health management of complex system.