A Health Indicator Construction Method based on Deep Belief Network for Remaining Useful Life Prediction

Ruihua Jiao*
School of Automation and Electrical Engineering,
University of Science and Technology Beijing
Beijing, PR China

Kaixiang Peng, Jie Dong, Kai Zhang, Chuan Zhang
School of Automation and Electrical Engineering,
University of Science and Technology Beijing
Beijing, PR China

***Abstract*—Remaining useful life (RUL) prediction is of great importance in a successful prognostics and health management system. The performance of RUL prediction is mainly decided by the development of an appropriate health indicator (HI), which can accurately indicate the degree of degradation of the equipment. Therefore, we proposed an unsupervised method for HI construction based on deep belief network (DBN) by using multisensory historical data. Firstly, DBN is employed to describe the hidden representation corresponding to the healthy state. With the running of the system, its performance will decrease over time and the corresponding potential characteristics tend to be different. The deviation degree of degraded state can be used to establish HI so as to estimate the RUL. Finally, a case study is conducted to validate the effectiveness of the new method, where it can be seen that the new approach achieves better performance compared to traditional methods.**

Keywords-deep belief network (DBN); Restricted Boltzmann machine (RBM); prognostics and health management; remaining useful life (RUL) prediction; health indicator (HI);

I. INTRODUCTION

The prognostics and health management (PHM) system can decide when to repair according to the current health state of the equipment before it has completely failed. Fatal failure may occur in engineering system due to aging or unexpected failures. In order to eliminate such risks, it is critical to make the reliable prediction of the remaining useful life (RUL). One of the challenges that need to be addressed is how to establish a health indicator (HI) that gives useful information for prognostic. Accordingly, the construction of HI which can reflect the degree of degradation plays a significant role in PHM technology. Furthermore, an appropriate HI is able to simplify the prognostic modeling and help to obtain more accurate RUL estimation results.

There exist a great many HI construction approaches that have already made great achievements. In [1], the drift-based degradation signals were used as the HI since the values of the gyros drift were related to the severity of the gyros degradation, and then predicted the RUL by a degradation model based on Wiener process. Liu et al. [2] used the capacity which can be measured by the integral of current over the time from fully charged state to fully discharged state as the HI for batteries. However, the degree of degradation is unable to be measured directly for some systems due to the following reasons. (1) The early damage often occurs on a microscopic scale and is intractable to measure. (2) For complex systems, there are many sensor signals that may have an impact on the health state of the system, and it is unreasonable to select one of them as the health indicator. Therefore, it is necessary to extract suitable HI from observed signals.

The HI construction approach can be classified into physics HI (PHI) and virtual HI (VHI) [3]. PHI is correlated with the physical characteristics of faults, which can be extracted from monitoring data by signal processing methods or statistical methods [4]. Gebraeel et al. [5] utilized average amplitude and harmonics as the HI of bearings. Malhi et al. [6] extracted root mean square and peak values as the HI, which is then utilized to predict the RUL of bearings. On the contrary, VHI is generally established by fusing multiple PHIs or directly extracting features from multi-sensor signals. Generally, the HI established in this manner generally has no physical meaning, but merely present the deteriorating trend of the equipment in a virtual way. In [7], the multiple linear regression method was used to compute HI by fusing multiple signals. In [8], the Euclidean distance between degradation state and the failure state was employed as HI. Wang et al. [9] calculated Mahalanobis distance between multiple features extracted from the health state and the degradation features as the HI.

Recently, deep learning methods have obtained remarkable achievements in many fields, including speech recognition [10], fault diagnosis [11], fault prognostics [12] etc. That is because the deep network architecture that multiple layers are stacked in can fully capture the representative information from raw sensor data. Deep belief network (DBN) [13], as a deep learning algorithm, is composed of a visible layer and many hidden layers which can learn the essential characteristics from raw input data. Motivated by the characteristics of deep network, an unsupervised method for constructing HI that are suitable for RUL prediction is proposed here. Firstly, HI is obtained from input raw measurement data by the unsupervised

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

DBN method. Following that, the constructed HI is employed as the input of the similarity-based prediction method for RUL estimation. Finally, the popular publicly available NASA C-MAPSS dataset [14] is adopted to demonstrate the superiority of the proposed approach compared with some other results on the same dataset.

The remaining parts is constructed as follows. Section II briefly introduces the detailed methodology of proposed method. In Section III the methods are applied to a common academic benchmark dataset, then the experimental results are analyzed. In Section IV, conclusions are drawn.

II. METHODOLOGY

The main contribution of this work is focused on proposing a novel unsupervised HI construction method, which is able to improve RUL estimation accuracy. In order to evaluate the prediction ability of the HI, a RUL prediction framework based on similarity-based prediction algorithm is presented as shown in Fig. 1.

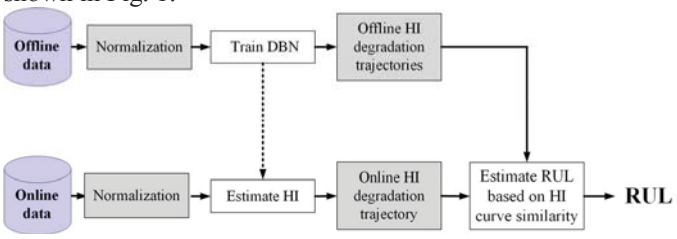


Figure 1. RUL prediction framework using unsupervised HI based on DBN.

A. Restrict Boltzmann Machine

Restricted Boltzmann machine (RBM) is one kind of generative stochastic artificial neural network, which can learn a probability distribution from its inputs. It has two layers of neural network which is composed of two units of visible and hidden elements. Every visible unit is linked with every hidden unit, and there is no relationship in any visible or hidden units, as illustrated in Fig. 2. The matrix W_{ij} is the connection weights of visible layer v_i and hidden layer h_j, and the bias are a_i and b_j respectively.

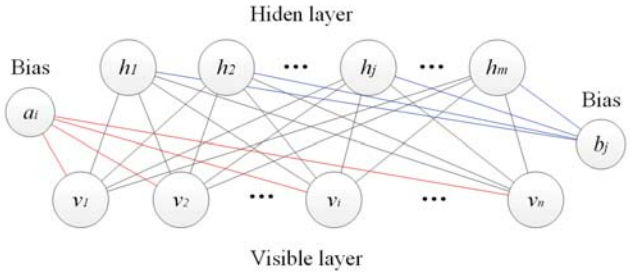


Figure 2. RBM structure diagram.

The weights and biases is able to be computed by maximizing the marginal distribution of v_i, which can be described as:

$$P(v)=\frac{1}{Z}\sum_{h}e^{-E(v,h;\theta)} \tag{1}$$

where $\theta=\left(a_i,b_j,w_{ij}\right)$, and Z represents the normalization factor, and it is computed as

$$Z=\sum_{v}\sum_{h}e^{-E(v,h;\theta)} \tag{2}$$

Thus, the energy function can be described as following:

$$E(v,h;\theta)=-\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}v_ih_j-\sum_{i=1}^{n}a_iv_i-\sum_{j=1}^{m}b_jh_j \tag{3}$$

The maximization of (1) can be determined by taking its partial log derivative with respect to its parameters a_i,b_j,w_{ij}:

$$-\frac{\partial\log\left(p(v)\right)}{\partial a_i}=\left\langle v_i\right\rangle_{\text{data}}-\left\langle v_i\right\rangle_{\text{model}} \tag{4}$$

$$-\frac{\partial\log\left(p(v)\right)}{\partial b_j}=\left\langle h_j\right\rangle_{\text{data}}-\left\langle h_j\right\rangle_{\text{model}} \tag{5}$$

$$-\frac{\partial\log\left(p(v)\right)}{\partial w_{ij}}=\left\langle v_ih_j\right\rangle_{\text{data}}-\left\langle v_ih_j\right\rangle_{\text{model}} \tag{6}$$

where the angle brackets indicate the expectation under the distribution designated by the subscript. Due to the expectation $\left\langle v_ih_j\right\rangle_{\text{model}}$ is difficult to calculate directly, we can approximate it by obtaining samples in the model distribution. Accordingly, the expectation can be computed by contrastive divergence (CD) algorithm [23], which has already become a common method of training RBMs. Thus, the parameters are derived by:

$$\begin{aligned}a_i^k&=a_i^k+\eta\left(\left\langle v_i^k\right\rangle_{data}-\left\langle v_i^k\right\rangle_G\right)\\&=P\left(v_i=1\,|\,h\right)-v_i^k\end{aligned} \tag{7}$$

$$\begin{aligned}b_j^k&=b_j^k+\eta\left(\left\langle h_j\right\rangle_{data}-\left\langle h_j^k\right\rangle_G\right)\\&=P\left(h_j=1\,|\,v\right)-P\left(h_j=1\,|\,v^k\right)\end{aligned} \tag{8}$$

$$\begin{aligned}w_{ij}^k&=w_{ij}^k+\eta\left(\left\langle v_i^kh_j\right\rangle_{data}-\left\langle v_i^kh_j\right\rangle_G\right)\\&=P\left(h_j=1\,|\,v\right)v_i-v_i^kP\left(h_j=1\,|\,v^k\right)\end{aligned} \tag{9}$$

where η denotes the learning rate, G is a full step of Gibbs sampling, and k is the sampling step parameter. And the conditional probability of v_i and h_j can be expressed as:

$$P\left(h_j=1\,|\,v\right)=g\left(b_j+\sum_{i=1}^{n}w_{ij}v_i\right) \tag{10}$$

$$P\left(v_i=1\,|\,h\right)=g\left(a_i+\sum_{j=1}^{m}w_{ij}h_j\right) \tag{11}$$

where $g(x)=1/(1+\exp(-x))$ is the logistic function, which is employed as a threshold function; n and m are the number of visible and hidden nodes respectively.

B. Deep Belief Network

The DBN is generally made up of multiple RBMs stacked on top of each other, and the output of the bottom one is also the input of the following RBM, as illustrated in Fig. 3. In addition, the DBN is trained in an unsupervised manner. Once input the data, DBN transmits the input data from the bottom layer to the top layer so as to extract the features, and CD algorithm is carried out to compute weights in a layer-wise pre-training manner. Afterwards, DBN is able to reconstruct the input signal by transferring the features from top to bottom. Meanwhile, the mean squared error (MSE) between the input data and the reconstructed data is utilized as the training error to fine-tune the obtained weights by using back propagation algorithm, as demonstrated in Fig. 4.

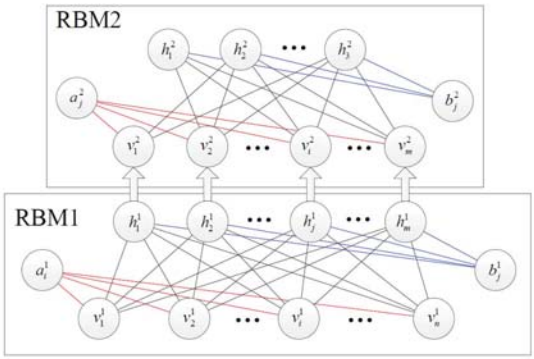


Figure 3. DBN with two hidden layers.

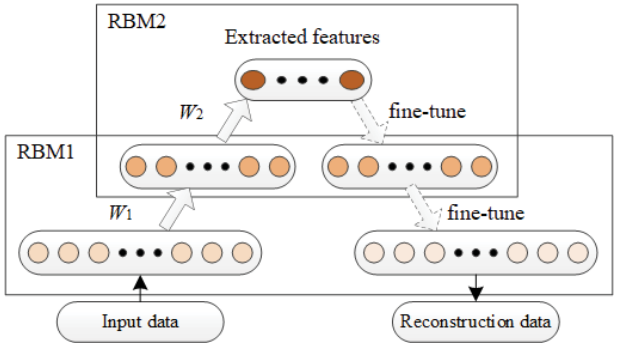


Figure 4. Unsupervised fine-tuning for DBN.

C. Obtaining HI using DBN

Here we introduce the HI construction approach by utilizing the previous described DBN. The core idea is that the hidden representation extracted from measurement data remains the underlying characteristic pattern of the input. Once the DBN obtains the characteristic patterns from the initial normal working data, the characteristics thus extracted from monitoring data can be applied to distinguish the healthy state and degraded state. Considering an input signal have m variables with N samples, the monitoring data need to do a data transformation by using zero-mean normalization so as to eliminate the influence of unit quantity between variables, which is expressed as:

$$X=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}=\begin{bmatrix} x_{11} & x_{12} & \cdots x_{1m} \\ x_{21} & x_{22} & \cdots x_{2m} \\ \vdots & \vdots & \ddots \vdots \\ x_{N1} & x_{N2} & \cdots x_{Nm} \end{bmatrix}\in R^{N\times m} \tag{12}$$

It is necessary to choose a suitable feature layer size so as to fully procure helpful characteristic from the input data and prevent overfitting. Once the network structure is decided, DBN will project the m-dimensional inputs variables x_i into n-dimensional characteristic space based on the multi-layer nonlinear network so as to obtain the characteristic variables z_i. Thus, a healthy characteristic set Z_{health}, can be procured from the normal working condition of the equipment. With the operation of the device, its health deteriorates over time and the corresponding potential characteristics tend to be different from those in Z_{health}. Accordingly, the HI at time i can be calculated by:

$$h_i=\min\left(\sqrt{\sum_{i=1}^{n}(z_i-z)^2}\right), z\in z_{health} \tag{13}$$

The HI curve obtained from the HI estimates at each time is expressed as $h=\{h_1,h_2,\cdots h_t\}$, and the HI values are generally assigned to have value from 0 and 1 [7,8], where 0 indicates failure and 1 indicates entirely healthy. However, HI defined in Eq. (13) follows the inverse definition, which means that it is low for normal health and high for poor health. This can be easily converted to the standard range of 0-1 by scaling procedure as following:

$$h_i'=\frac{h_i-h_m}{h_M-h_m} \tag{14}$$

where h_i and $h_i^{'}$ are the HI value before and after normalization respectively; h_m is the minimum HI value while h_M is the maximum HI value.

D. RUL estimation using HI

The similarity-based prediction method is employed here to estimate the RUL. At first, the HIs of training data are calculated by using DBN to generate a training library. Afterwards, a small amount of HI degradation trajectories in the training library, which are similar to the current HI degradation trajectory, are used to predict the RUL. This can done by measuring the similarity between the HI degradation trajectory of a test set and the HI degradation trajectories in the training library. The similarity measure should consider the tendency of entire history and assign more importance weight to the late time. It is because that there is little difference in the initial state of each equipment, and with the increasing of

operating time, the HI tends to be different due to the different degradation rate of degradation.

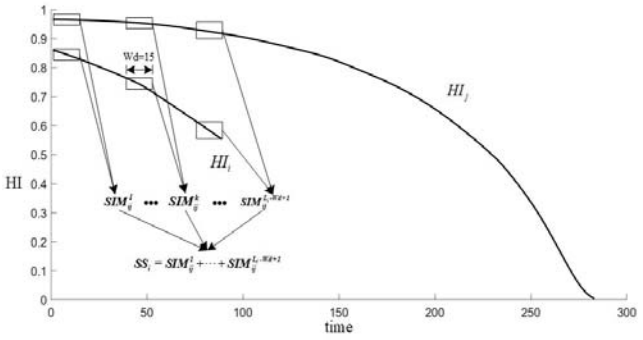


Figure 5. Similarity measurement.

As illustrated in Fig. 5, HI_i and HI_j are the HI of the test set and the training set, respectively. The HI trajectories are divided into windows of size W_d. At the time of testing, each window is scanned over the training HI degradation trajectory, and the Euclidean distance of the test window and training window can be calculated by

$$D_{ij} = \sum_{s=1}^{L_i - Wd + 1} \left(y_{train} - y_{test} \right)^2 \tag{15}$$

Then each window is assigned different weights by a way that gives large weights to the late windows, and the similarity can be calculate by

$$SIM_{ij} = \frac{1}{D_{ij}} \bullet \exp\left(\frac{(s \bullet \lambda)}{L_i - Wd + 1} \right) \tag{16}$$

where L_i denotes the total length of time of the test data. The additional factor λ can control the size of the weight, and the larger is the greater the weight, if λ equals 0, the weight of each window is the same. At the same time, a similarity score can be given to the estimation:

$$SS_i = \sum_{s=1}^{L_i - Wd + 1} SIM_{ij} \tag{17}$$

The final estimated RUL can be predicted by weighted sum of the RULs of the selected k most similar training units with the highest similarity scores.

$$eRUL = \sum_{s=1}^{k} w_s \bullet RUL_s \tag{18}$$

$$w_i = \frac{SS_s}{\sum_{s=1}^{k} SS_s} \tag{19}$$

The weights w_i can be assigned based on the similarity score, and k is the number of selected neighbors.

III. APPLICATION AND RESULTS

The proposed approach was carried out in a turbofan engine degenerate dataset produced by C-MAPSS [16], which has a wide application in prognostic research. The schematic diagram of the turbofan engine is shown in Fig. 6.

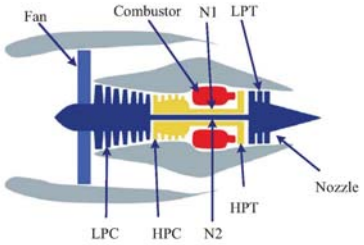


Figure 6. Simplified diagram of the engine

A. Experiment data

On the basis of the C-MAPSS model, four datasets were obtained. Every training engine is run until it is fails, while every test engine is stopped at a random point before failure. Besides, the obtained measurement signals contain some random noise and the exact failure threshold value is not clearly provided to the users. More detailed description about the engine experiment can be found in reference [14].

In this work, the first dataset was used. It contains a total of 26 signals. 21 of them represent the measurement data, 3 others denote the operational settings, and the last 2 records are the ID of engines and number of cycles. The dataset was collected from the engine under a single operating condition and contained only a single failure mode. The task is to predict the RUL of these 100 test engines.

B. Evaluation metrics

In order to demonstrate the superiority of the proposed new method, two common evaluation metrics are adopted here:

1) Score: the score gives an overview of the overall error, and it was once utilized in PHM 2008 data challenge [14],

$$S = \sum_{n=1}^{N} s_n, s_n = \begin{cases} \exp\left(-\frac{e}{10} \right) - 1, & e \le 0 \\ \exp\left(\frac{e}{13} \right) - 1, & e > 0 \end{cases} \tag{20}$$

where S denotes the score, e is the prediction error which can be computed as $e = actual\ RUL - predicted\ RUL$, and N is the number of engines.

2) Performance: the performance denotes the percentage of correct predictions [17], which can reflect the accuracy of the algorithm. The prediction result is considered to be correct if the prediction error falls into the interval [−10, 13].

C. Model learning and parameter selection

In order to procure Z_{health}, the first 5% of measurement data in training set is considered as health state. 80 engines of

training set are randomly chosen to train the DBN model and the other 20 engines are regarded as validation set. On account of the limited size of the training data, the number of neural network layers for each DBN is set to 3, which means the DBN has two RBMs. Besides, the number of input nodes is equal to the size of measured variables, and the number of hidden nodes is set to 50. 12 is chosen as the number of characteristic variables based on the relationship between the number of features and MES, which is shown in Fig. 7.

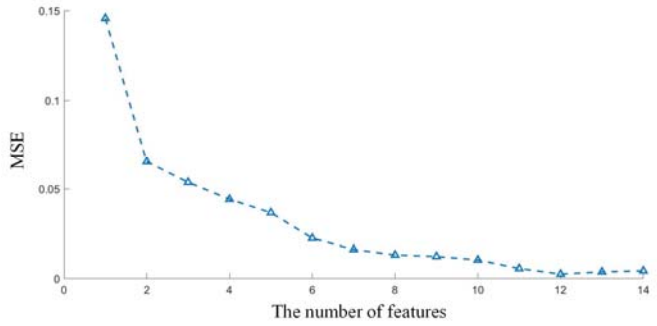


Figure 7. The relationship between the number of features and MES.

D. Results and Comparison

The degenerate dataset contains 21 measurement signals. However, the values of some signals are constant during the whole life cycle of the engine, which is useless to estimate the RUL of engines. Therefore, 14 sensors with indices 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21 were selected as the input variables. Following that, the signals were filtered by using the moving average method with a slide window of 50. Finally, the Eq. (13) and (14) is applied to derive HI, and one of the results is depicted in Fig. 8. As can be seen that the constructed HI continually decreases with the increases of the cycles. This demonstrates that the established HI is able to be considered as an indicator of health for equipment.

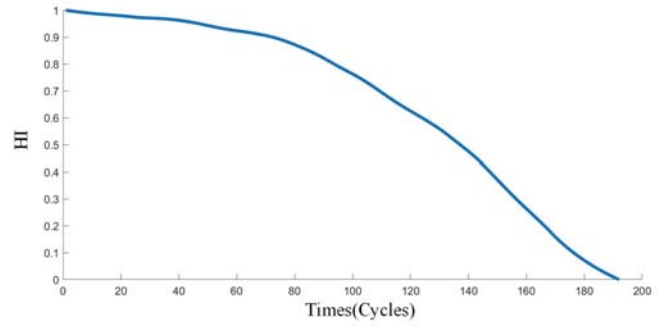
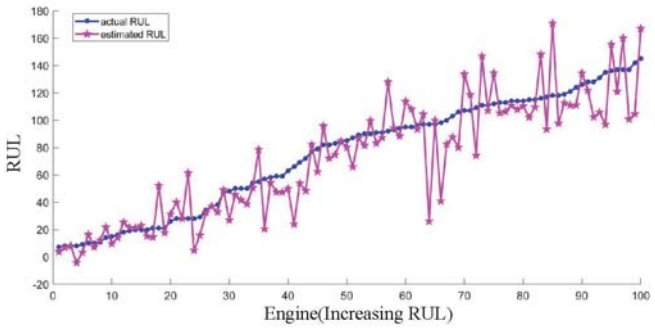
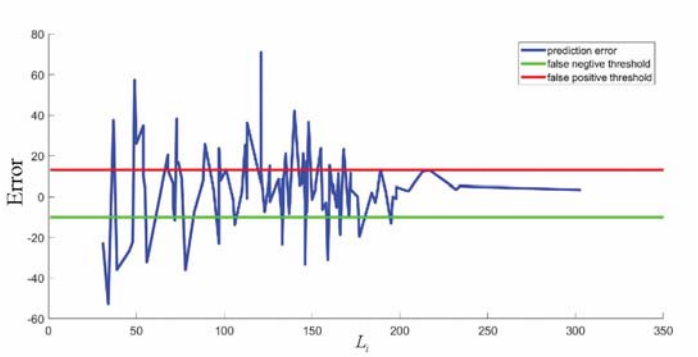


Figure 8. HI constructed by the proposed method.

For the RUL estimation, the following parameter values were used: $W_d = 15$, $\lambda = 2$, $k = 5$. These values were selected after a large number of experiments. After that, the RUL of each test engine can be obtained, and the results are shown in Fig. 9.



(a) RUL estimated by proposed method



(b) Error evolution between actual and estimated RUL

Figure 9. RUL prediction results.

Fig. 9a shows the actual RUL and the estimated RUL, as can be seen that the prediction accuracy increases gradually with the decreasing of actual RUL. Fig. 9b describes the evolution of the error between actual RUL and estimated RUL. It can be seen that the prediction error and uncertainty are very high when the running time L_i is short, but with the L_i increases, the prediction error and uncertainty decreases, which is also consistent with Fig. 9a. That is because it is intractable to accurately estimate the RUL of test unit during the early phase with the low availability of historic data.

In order to show the superiority of the proposed approach, as described in [18], the conventional similarity-based method [7] and the PCA-based method [8] were used for the comparation. For the conventional method, the HI is established by multiple linear regression. For the PCA-based method, the HI is established by PCA algorithm. Table 1 summarizes the evaluation results of the different approaches with respect to the score and performance.

TABLE I. APPROACH COMPARISON

Approach	*Score*	*Performance*
Conventional method [7]	1389.26	44
PCA-based method [8]	771.21	N/A
Proposed approach	532.34	56

As can be observed from Table I that the proposed method achieved best score, and the performance is also better than conventional method. Thus, the proposed HI construction

method based on DBN is very suitable for RUL estimation and has some advantages in improving the prediction accuracy. The reason is that the conventional approach and the PCA-based approach can only capture the linear relationship between input data and HI, while DBN based on deep nonlinear network architecture can capture more representative features from input data.

IV. CONCLUSION

This work present an unsupervised HI construction method based on DBN, and the HI is further applied to RUL prediction. Firstly, HI was extracted from input raw measurement data by DBN. Following that, the constructed HIs were used as the input of similarity-based prediction algorithm for RUL prediction. Finally, the proposed approach was conducted on turbofan engine degenerate dataset proposed by NASA. Compared with some existed methods, the DBN-based HI construction approach is able to extract more representative characteristic so as to improve the RUL prediction performance.

In the future, we would like to extend the proposed approach to the application of real-world scenarios. On the other hand, the proposed HI can be incorporated with some other prediction methods to obtain the confidence interval of prediction result, which is the very useful information for health management.

ACKNOWLEDGMENT

This paper was supported by the National Key R&D Program of China (No.2017YFB0306403), the Natural Science Foundation of China (NSFC) under Grants (61873024, 61773053), and the Fundamental Research Funds for the China Central Universities of USTB (FRF-BD-18-002A), PR China.

REFERENCES

[1] X.S. Si, W. Wang, C.H. Hu, and D.H. Zhou, "Estimating remaining useful life with three-source variability in degradation modeling," IEEE Trans. Rel., vol. 63, no. 1, pp. 167–190, Mar. 2014.

[2] Z. Liu, G. Sun, S. Bu, J. Han, X. Tang, and M. Pecht, "Particle learning framework for estimating the remaining useful life of lithium-ion batteries," IEEE Trans. Instrum. Meas., vol. 66, no. 1, pp. 280–293, Feb. 2017.

[3] C.H. Hu, B.D. Youn, and P. Wang, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," Reliab. Eng. Syst. Saf., vol. 103, no. 3, pp. 120–135, Jul. 2012.

[4] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: a systematic review from data acquisition to rul prediction," Mech. Syst. Signal Proc., vol. 104, pp. 799–834, May. 2018.

[5] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," IEEE Trans. Ind. Electron., vol. 51, no. 3, pp. 694–700, Jun. 2004.

[6] A. Malhi, R. Yan, and R.X. Gao, "Prognosis of Defect Propagation Based on Recurrent Neural Networks," IEEE Trans. Instrum. Meas., vol. 60, no. 3, pp.703–711, Mar. 2011.

[7] T. Wang, J. Yu, D. Siegel, and J. Lee, " A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems," in 2008 Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 2008, pp.1–6.

[8] L.S. Khanh, F. Mitra, B. Anne, L. Eric, and L. Benoit, "Remaining useful life estimation based on stochastic deterioration models: a comparative study," Reliab. Eng. Syst. Saf., vol. 112, no. 4, pp. 165–175, Apr. 2013.

[9] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," IEEE Trans. Industr. Inf, vol. 12, pp. 924–932. Jun. 2016.

[10] G. Hinton, L. Deng, D. Yu, G. Dahl, A.R. Mohamed, and N. Jaitly, " Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, Nov. 2012.

[11] J. Dai, H. Song, G. Sheng, and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," IEEE Trns. Dielectr. Electr. Insul., vol. 24, no. 5, pp. 2828–2835, Oct. 2017.

[12] K. Peng, R. Jiao, J. Dong, Y. Pi, "A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter," Neurocomputing, vol. 361, pp. 19-28, Oct 2019.

[13] G.E. Hinton, S. Osindero, and Y.W. Teh, "A fast learning algorithm for deep belief nets," Neural Comput., vol. 18, no. 7, pp. 1527–1554, Jul. 2014.

[14] A. Saxena, K Goebel, D. Simon,and N. Eklund, "Damage propaga-tion modeling for air- craft engine run-to-failure simulation," in 2008 Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 2008, pp.1–6.

[15] A. Fischer, C. Igel, "An Introduction to Restricted Boltzmann Machines," in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Berlin, Heidelberg, 2012, pp. 1436.

[16] D.K. Frederick, J.A. Decastro, and J.S. Litt, "User's guide for the commercial modular aero-propulsion system simulation (c-mapss). Technical Report, National Aeronautics and Space Administration (NASA)," Glenn Research Center, Cleveland, Ohio, USA, 2007.

[17] R. Khelif, B.M. Chebel, S. Malinowski, E. Laajili, and N. Zerhouni, " Direct Remaining Useful Life Estimation Based on Support Vector Regression," IEEE Trans. Ind. Electron., vol. 64, no. 3, pp. 2276–2285, Mar. 2017.

[18] Z.M. Liang, J.M. Gao, H.Q. Jiang, X. Gao, Z.Y. Gao, and R.X. Wang, "A similarity-based method for remaining useful life prediction based on operational reliability," Appl. Intell., vol. 48, pp. 2983–2995, Sep. 2018.