Data-Driven Prognosis Method using Hybrid Deep Recurrent Neural Network

Min Xia[a], Xi Zheng[b], Muhammad Imran[c,*], Muhammad Shoaib[c]

[a] *Department of Engineering, Lancaster University, Lancaster, UK*
[b] *Yangtze River Delta Research Institute, Northwestern Polytechnical University, Taicang, Jiangsu, China*
[c] *College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia*

Abstract

Prognostics and health management (PHM) has attracted increasing attention in modern manufacturing systems to achieve accurate predictive maintenance that reduces production downtime and enhances system safety. Remaining useful life (RUL) prediction plays a crucial role in PHM by providing direct evidence for a cost-effective maintenance decision. With the advances in sensing and communication technologies, data-driven approaches have achieved remarkable progress in machine prognostics. This paper develops a novel data-driven approach to precisely estimate the remaining useful life of machines using a hybrid deep recurrent neural network (RNN). The long short-term memory (LSTM) layers and classical neural networks are combined in the deep structure to capture the temporal information from the sequential data. The sequential sensory data from multiple sensors data can be fused and directly used as input of the model. The extraction of handcrafted features that relies heavily on prior knowledge and domain expertise as required by traditional approaches is avoided. The dropout technique and decaying learning rate are adopted in the training process of the hybrid deep RNN structure to increase the learning efficiency. A comprehensive experimental study on a widely used prognosis dataset is carried out to show the outstanding effectiveness and superior performance of the proposed approach in RUL prediction.

*Corresponding author

Keywords: Long Short-Term Memory, Prognostics, Recurrent Neural Network, Remaining Useful Life Prediction

1. Introduction

The reliability and availability of the manufacturing systems are crucial in maintaining the system productivity and safety [1]. Great efforts have been allocated in the investigation of maintenance policies of the machines and equipment [2]. The development of smart manufacturing has brought even higher requirements regarding the accuracy and efficiency of the maintenance actions [3]. The traditional corrective maintenance or preventive maintenance strategy can barely satisfy the new requirements [4]. The corrective maintenance leads to significant interruption of the production due to the occurrence of the failure. The preventive maintenance can cause a notable increase in maintenance costs with unnecessary maintenance actions. On the other hand, the prognostic and health management (PHM) has shown promising capabilities in providing precise maintenance decisions through estimating and predicting the conditions of the running machinery [5]. It can prevent both system failures and unnecessary maintenance to reduce the overall cost [6].

Remaining useful life (RUL) estimation is the core task in prognostics. The current approaches for RUL prediction can be categorized into three classes: model-based methods, data-driven methods, and hybrid approaches [7]. Model-based RUL prediction methods can be precise if the physical model of the degradation can be accurately derived. However, with the increased complexity of modern machines and components, the accurate failure models are difficult to build. On the other hand, with the advances in sensor technology, communication technology, especially the industrial internet of things [8], condition monitoring and intelligent analysis system as described in Fig. 1 has been developed for various industrial machinery such as wind turbine [9], high speed railway [10], airplane [11], etc. With the fast development and the increasing applications of IIOT in advanced manufacturing and smart factories, a large amount of

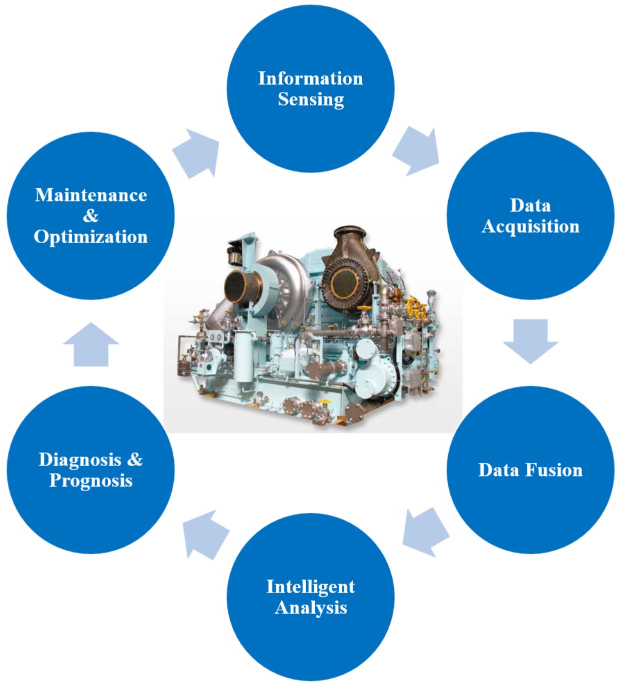

Figure 1: Data-driven machine condition monitoring system.

monitoring and operation data of the machines are available in further analysis for fault diagnosis and prognosis. Hybrid approaches aim at combining model-based and data-driven methods to overcome the limitations of the individual methods [12]. However, it is still based on the availability of the analytical model [13]. Therefore, data-driven approaches have become more effective and preferable in many applications [14].

To discover the relation between the sensory data and the RUL of the monitored machines, various data-driven approaches incorporating machine learning techniques such as artificial neural networks (ANNs), support vector regression (SVR), and random forest (RF) have been developed. Tian [15] developed an ANN-based RUL prediction with two hidden layers. The vibration signals were used to the degradation process of the bearings. Features were first extracted from the vibration signals. Then Weibull failure rate function was utilized to fit the data. Their approach outperformed the model with one hidden layer because of the improved capability in nonlinearity modeling. Nieto *et al.* [16] proposed an RUL prediction method based on hybrid particle swarm optimization (PSO) and SVR. In their study, the PSO technique was used to optimize the hyperparameters corresponding to the best SVR model for the RUL prediction. Khlif *et al.* [17] developed a RUL estimation approach based on SVR. A direct relation between sensor values or health indicators was modeled using SVR. Wu *et al.* [18] trained a predictive model based on RFs that achieved an accurate prediction of tool wear in milling processes. They achieved better prediction performance than traditional approaches using ANN and SVR.

The traditional machine learning-based approaches usually contain the step of handcrafted feature extraction from the sensory data. This step is labor extensive as adequate prior knowledge and domain expertise are needed. In recent years, deep learning techniques have made extraordinary progress in many applications due to its powerful capability of end-to-end learning. The deep structure can autonomously learn the representative features through the training process. Different types of deep neural networks have been successfully applied in various classification and regression tasks e.g., computer vision [19], natural language processing (NLP) [20], bioinformatics [21], etc. Recently, researchers in the prognosis area have also applied deep learning methods to machine fault diagnostics and prognostics [22, 23, 24, 25]. Zhang *et al.* [26] developed a prognosis approach based on multiobjective deep belief networks (DBNs). Multiple DBNs were evolved simultaneously subject to accuracy and diversity as two conflicting objectives. RUL prediction can be achieved by the final model with the combination of the evolved DBNs. Xia *et al.* [27] proposed an RUL prediction method by introducing a hierarchical deep neural network (DNN). Different health stages were first classified using DNN. RUL prediction models on different health stages were constructed with several shallow neural networks. However, the temporal correlation of the degradation data was not fully utilized in this approach. Also, the DNN approaches often have issues of high computational cost and overfitting due to the huge number of model parameters.

The recurrent neural network (RNN) has been successfully applied in many sequential data classification or regression tasks such as NLP, music generation, speech recognition, etc., due to its capability of modeling time dependency [28]. The shared weights of RNN can also decrease the requirement of computational power as well as the chance of overfitting. Researchers have also tried RNN in prognostics [29]. However, the standard RNN comes with the problem of gradient vanishing during the training process. Therefore, it is difficult to model dependence over a long time that limits the application of standard RNN in prognosis. A variant structure of RNN, the long short-term memory (LSTM) is designed to discover time dependence of a long time by incorporating several gate functions [30]. In order to capture the temporal information from the degradation process and to utilize the raw sensory data directly, in this paper, we develop a data-driven prognosis model using a hybrid network with deep LSTM and classical neural networks. It uses the raw sensory data to achieve the RUL prediction. The main contributions of this paper are listed as follows.

(1) A prognosis approach using the hybrid deep LSTM network is proposed. On one hand, temporal correlation in the sequential sensory data is captured by the deep LSTM network. On the other hand, the classical neural network layers enhance the capability of modeling nonlinearity between the different sources of data. The hybrid approach developed shows improved performance over existing approaches in RUL prediction.

(2) The proposed method can work on the raw sensory data directly without any handcrafted feature extraction. The deep structure can learn the features automatically during the training process. It can be applied where few prior knowledge or domain expertise is available.

(3) The typical problem of deep neural network method, overfitting, is prevented by incorporating the dropout technique. Also, the training efficiency is improved by using the decaying learning rate.

The rest of the paper is organized as below. Section II introduces the knowledge of RNN including its structure and training scheme. Also, the algorithm of LSTM is discussed with a detailed presentation of its internal operations. Section III presents the proposed RUL prediction approach using hybrid deep

LSTM. The overall procedure is first given followed by the detailed demonstration of the model structure. The experimental study is discussed in Section IV. The benchmark prognosis dataset, the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset is used to test the proposed approach. The proposed method shows excellent performance in RUL prediction. Section V makes a conclusion of this work and presents possible future work.

2. Background

2.1. Recurrent Neural Network

The recurrent neural network (RNN) is a deep neural network structure with a loop inside [31]. The recurrent connection can store the information from the previous inputs within the network hidden states. Therefore, RNN can effectively model sequential data with temporal dependence. A typical structure of the RNN is shown in Fig. 2.

The sequential data $\mathbf{x} \in \mathbb{R}^{p \times T}$ is inputted into the RNN model. Data at each of the T time steps contains p elements. The output is $\mathbf{y}$. $s_t \in \mathbb{R}^n$ is the intermediate hidden state at time step t that is dependent on the current input x_t and the hidden state of the previous time step s_{t-1}. n is the number of hidden units of the RNN network. The detailed procedure of obtaining the hidden state can be seen in Fig. 3. The hidden state can be seen as the memory of the network that stores the information captured by all the previous steps. The hidden state value at time step t can be calculated using the following equation.

$$s_t = f(Vx_t + Ws_{t-1} + b_s) \tag{1}$$

Here, f is the activation function. $V \in \mathbb{R}^{n \times p}$ is a weighting matrix that operates on the original input forming the input to the hidden state. $W \in \mathbb{R}^{n \times n}$ is a weighing matrix between hidden states that controls the memory of the network. $b_s \in \mathbb{R}^{n \times 1}$ a bias vector. Therefore, the current hidden state depends

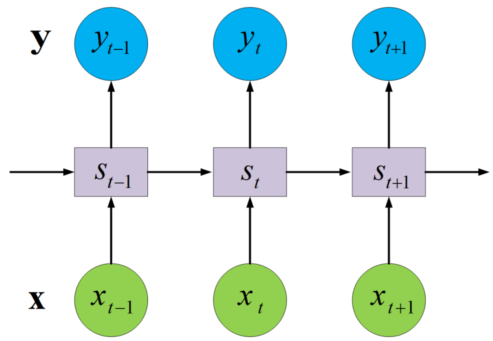

Figure 2: A typical structure of the RNN.

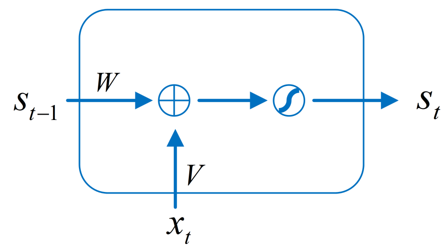

Figure 3: Operations in the RNN model.

on the current input and the previous hidden state containing the memory of the previous information.

The training of a RNN is similar to the training of the traditional fully connected neural network using backpropagation. Here, unlike the traditional neural network, the RNN model shares the same parameters across all time steps. Therefore, the gradient depends on not only the current time step but also all the previous time steps. This backpropagation method is named backpropagation through time (BPTT). However, the basic RNN model has a problem of vanishing gradient. When the model is trained using BPTT, the gradient of the error with respect to previous inputs can vanish quickly. Therefore, the basic RNN can hardly model long term dependency because of this problem.

2.2. Long Short-Term Memory

The long short-term memory network (LSTM) is a variant version of RNN. LSTM replaces the simple hidden state calculation with several gate functions

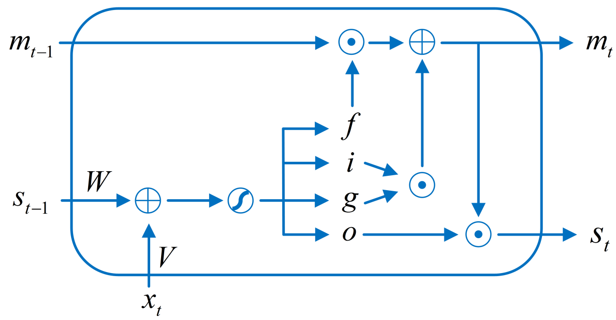

Figure 4: Operations in the LSTM model.

[32]. This mechanism permits the LSTM to capture long term dependency in the temporal sequential data. The operations in the LSTM model are shown in Fig. 4.

Compared with classical RNN, the LSTM network introduces a new flow, the cell state $m_t \in \mathbb{R}^n$ shown at the top of the cell structure in Fig.4. LSTM has the capability of adding or removing information to the cell state. The cell state maintains the memory of the LSTM. The three gates that control the information flow in LSTM include the input gate $i_t \in \mathbb{R}^n$, the forget gate $f_t \in \mathbb{R}^n$, and the output gate $o_t \in \mathbb{R}^n$. The input gate adjusts the level of information from the current input x_t and the previous hidden state s_{t-1} will be inputted into the current state. The forget gate controls how much information from the previous cell state m_{t-1} will be maintained. The output gate controls how much information will be passed to the current hidden state s_t. The operations of these gates are as follows:

$$i_t = \sigma(V_i x_t + W_i s_{t-1} + b_i) \tag{2}$$

$$f_t = \sigma(V_f x_t + W_f s_{t-1} + b_f) \tag{3}$$

$$o_t = \sigma(V_o x_t + W_o s_{t-1} + b_o) \tag{4}$$

where the parameters V_i, V_f, $V_o \in \mathbb{R}^{n\times p}$; W_i, W_f, $W_o \in \mathbb{R}^{n\times n}$; b_i, b_f, $b_o \in \mathbb{R}^{n\times 1}$. σ is the sigmoid activation function.

Then, the cell state and hidden state are obtained by using the following equations.

$$g_t = \tanh(V_m x_t + W_m s_{t-1} + b_m) \tag{5}$$

$$m_t = f_t \circ m_{t-1} + i_t \circ g_t \tag{6}$$

$$s_t = o_t \circ \tanh(m_t) \tag{7}$$

where $V_m \in \mathbb{R}^{n\times p}$; $W_m \in \mathbb{R}^{n\times n}$; $b_m \in \mathbb{R}^{n\times 1}$. tanh is the hyperbolic tangent activation function. $\circ$ is element-wise multiplication.

The training of LSTM can be done using the BPTT by minimizing the objective function on a set of training sequences. The gradients of weights and biases can be calculated at all time steps. Then, with the classical optimization algorithms e.g. Stochastic Gradient Descent, Adam or RMSprop, the optimal parameters can be obtained.

3. Methodology

In this section, the proposed RUL estimation approach based on hybrid deep LSTM is introduced. The deep network is composed of both LSTM layers that capture the temporal information and the classical neural network layers that bring the enhanced capability of modeling non-linearity of the degradation process. The developed hybrid deep neural network model can precisely predict the RUL of the monitored machinery after training the model with the training dataset. The collected data is used as input of the model without any manual feature extraction.

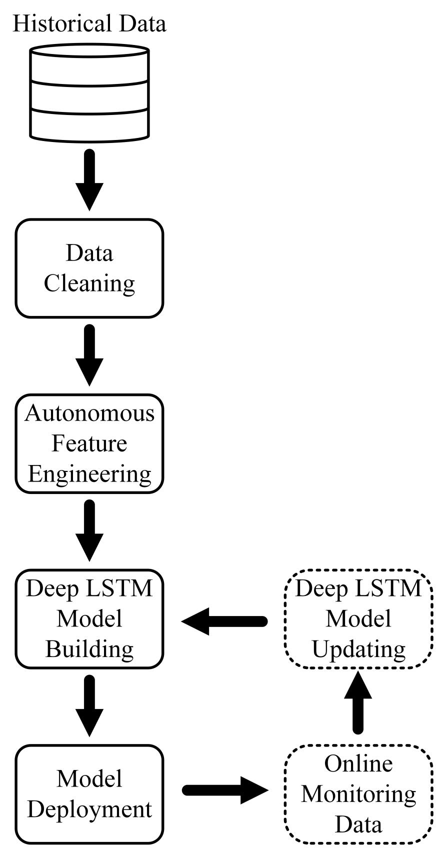

Figure 5: Flowchart of the proposed approach for RUL estimation.

3.1. Proposed approach

The proposed RUL prediction method aims at providing accurate RUL prediction through autonomous feature extraction from the historical and online monitoring data. The overall flowchart of the developed approach for RUL estimation is shown in Fig. 5. The four main steps of the proposed approach include data acquisition, data cleaning, deep LSTM model building, and online RUL prediction. The detailed procedure of each step is explained as follows.

(1) Data acquisition

In this step, different types of sensors, e.g. accelerometer, acoustic emission sensor, thermometer, microphone, etc. can be used to acquire the run to failure signals from the same type of machinery. The selection of sensors is based on the domain knowledge on which sensory data is closely correlated to the degradation process of the monitored machinery. A certain amount of run to failure data of the monitored machinery is needed. The acquired sensory data will be used to train the RUL prediction model.

(2) Data cleaning

The data collected in real applications can be noisy and may contain some errors or missing values. Also, the scale of data from different types of sensors can vary significantly. Therefore, data cleaning is necessary and important for further processing. Here, the z-score normalization is used to normalize the data so that the collected data have zero mean and unit variance. The normalization can be achieved using the following equation.

$$x_i{}' = \frac{x_i - \mu_i}{\sigma_i} \tag{8}$$

Here, $x_i{}'$ is the normalized data. x_i is the measurement data from the i-th sensor. μ_i is the mean value of the measurement. σ_i is the standard deviation.

(3) Deep LSTM model building

In this step, a hybrid deep neural network with LSTM layers and classic neural network layers are constructed. We take advantage of the LSTM with its nature of modeling sequential data and the capability of modeling non-linearity from classic fully connected layers. With the deep structures, the network can use the raw sensory data to extract representative features automatically through the training process. Therefore, manual feature extraction, which relied heavily on the expertise and prior knowledge in traditional approaches, can be avoided. Also, sensor fusion can be achieved with the proposed model by fusing the data directly at the data level. After data cleaning, the data collected from multiple sources at each time cycle are treated as different features of the input to the LSTM layer. The fused input can increase the accuracy of the RUL prediction compared with using data from one source. The historical data of the degradation process will be used to train the model. In this paper, Adam optimizer is used due to the better performance in deep learning applications compared with other optimizers such as the traditional stochastic gradient descent method [33]. To increase learning efficiency, the decaying learning rate is used here by adjusting the learning rate through the training process. The learning rate is reduced by a certain portion after every n epochs using the following equation.

$$R_i = R_0 \times \rho^i \tag{9}$$

where $i \in [1, 2, ..., N/n)$ and N is the number of total epochs. R_0 is the initial learning rate. R_i is the learning rate for the training after $n \times i$ epochs and before $n \times (i+1)$ epochs. $\rho \in (0, 1)$ is a dropping factor. The learning rate is decaying during the training process that can give better training results.

(4) Online RUL prediction

After the training of the deep LSTM model, the RUL prediction for the running machinery is carried out with the corresponding sensory data. The estimated RUL can provide significant information for the decision making of proper maintenance actions. In our approach, the online monitoring data can be used to further improve the deep LSTM model by fine-tuning the model with newly collected data. The RUL prediction model can then be continuously enhanced.

3.2. Deep LSTM-based RUL prediction

In our proposed approach, a hybrid deep structure is constructed to model the degradation process of the rotating machinery. With the capability of capturing temporal dependency, LSTM layers are used. After each layer of the LSTM, a fully connected layer is added. The hybrid building block can take advantage of the two types of layers. The number of the building blocks can be determined by starting with a small number until the network starts to overfit the training data which means the testing accuracy starts to decrease with a more complex model. Finally, a fully connected and regression layer is added to

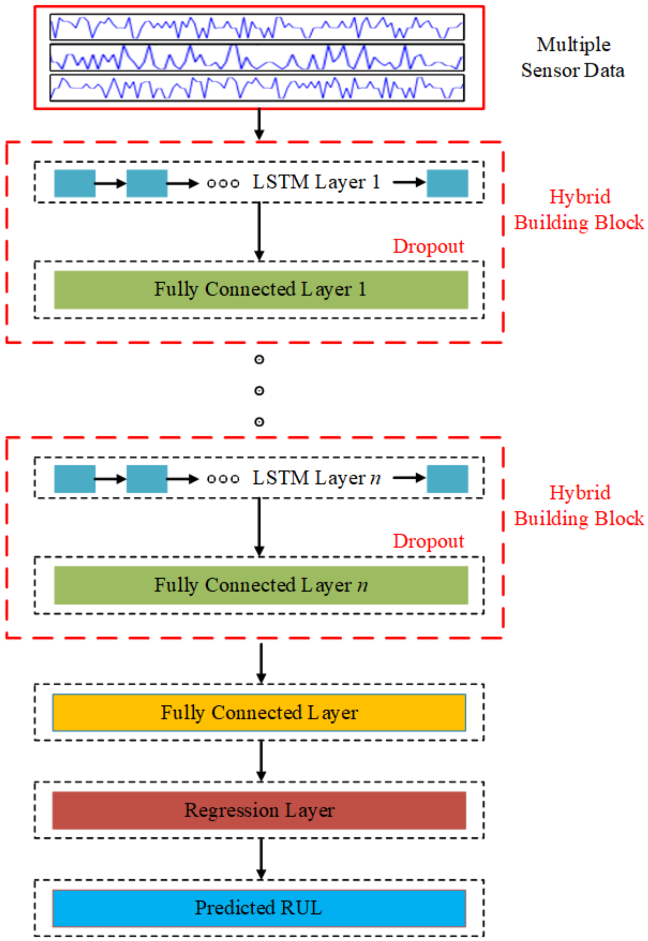

Figure 6: Network Structure of the Deep LSTM.

predict the RUL of the rotating machinery. The detailed structure of the network is shown in Fig.6. The parameters to determine for the network structure include the number of hybrid building blocks, the number of hidden units in LSTM layers, and the number of units in the fully connected layers. The grid search strategy can be used to determine these parameters.

Overfitting can be a problem in the training of many DNN models that can cause unsatisfactory performance. In this paper, the dropout technique is utilized to prevent overfitting. In each fully connected layer in the hybrid building block, dropout is triggered during the training process. A certain amount of neurons of the layer are temporarily removed to form a reduced network. It is achieved by assigning zero value to the randomly selected feature map elements. In the testing step after obtaining the trained model, the dropout mechanism is switched off. Dropout has been successfully applied to decrease the chance of overfitting.

4. Experimental Study

In this section, RUL prediction using the proposed hybrid deep LSTM approach is evaluated using a widely tested benchmark dataset. The results of RUL prediction using the proposed method are compared with those obtained by existing state-of-art approaches. The experiment study demonstrates the effectiveness of the proposed method.

4.1. Data Description

The NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset is used in this paper with simulated turbofan engine degradation data [34]. It contains four sub-datasets with different operating conditions and fault modes. Each sub-dataset includes training dataset and testing dataset. The training dataset is composed of run-to-failure sequential data collected from 21 sensors. The detailed information about the sensors can be referred to in [34]. The engine operates normally at the beginning with certain degrees of initial wear. The sensors record the data of the engine until the fault develops to a system failure. In the test dataset, the sensory data of the system prior to the system failure are recorded. The task is to estimate the RUL of the engine in the testing dataset. Therefore, in the testing dataset, the actual RUL of each data sample is provided to check the result of the proposed method. Table 1 lists the details of the dataset. About 50% of the total samples are designed for training in this dataset. The rest of the samples are used for testing.

Table 1: Details of the C-MAPSS dataset

Sub-data det	FD001	FD002	FD003	FD004
Training sample	100	260	100	249
Testing sample	100	259	100	248
Operation condition	1	6	1	6
Fault mode	1	1	2	2

4.2. Experiment Environment and Network Structure

The experiment is carried using a computer with the following configuration: Intel Core i7-4790K(4.00GHz) CPU, 16GB RAM, NVIDIA GeForce GTX 1080 GPU, Microsoft Windows 7 Enterprise. Matlab 2018a is used as the programming tool to code the proposed algorithms.

In this paper, the structure of the hybrid deep LSTM network is determined by the method of grid search. The hyper parameters that need to be figured out are the number of hybrid building blocks, the number of hidden units in the LSTM layer, and the number of units in the fully connected layer. In this experiment, to avoid a huge grid search space, we assign the same number of hidden units in different LSTM layers. We vary the number of hybrid building blocks from 1 to 6, the number of hidden units in LSTM layers from 16, 32, 64 to 128, the number of units in the fully connected layer from 16, 32, 64 to 128. Ten-fold cross validation is used to train each structure using the training dataset of the first sub-dataset. The average training time of the most complicated network structure, six building blocks with 128 hidden units in LSTM layers and 128 units in a fully connected layer, is 336.5 seconds. RUL prediction of the test data is then performed. The root mean square error (RMSE) of the RUL prediction, as given in Equation 10, is selected as the metric to evaluate the performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{RUL}_i - RUL_i)^2} \tag{10}$$

Table 2: Experiment results of the RUL prediction

Sub-data det	RMSE Mean	RMSE STD
FD001	14.57	1.04
FD002	23.20	1.73
FD003	14.92	2.26
FD004	28.72	1.26

where n is the number of test samples. $\widehat{RUL}_i$ is the estimated RUL of the i-th sample and RUL_i is the actual RUL.

Among all the combinations of the hyperparameters, the network structure with three hybrid building blocks with 128 hidden units in LSTM layers and 128 nodes in the fully connected layers achieved the best results. The corresponding average training time of this model is 125.7 seconds. Therefore, in this experiment, this structure is selected and used to test all the sub-datasets.

4.3. RUL Prediction Using the Proposed Method

In this section, the evaluation of the performance of the proposed RUL prediction method is presented. The network structure determined in the previous section is tested on all the testing data from the four sub-datasets. The final results are compared with the results using the existing data-driven approaches.

In this study, 10 trials are carried out using the proposed approach. To reduce randomness, the averaged result is used to evaluate the performance. Table 2 lists the mean value and standard division (STD) of the RMSE on the four sub-datasets. The RUL prediction using the proposed method achieves satisfactory results on both FD001 and FD003 sub-datasets. The RMSE for the two cases are 14.57 ± 1.04 and 14.92 ± 1.26. The prediction errors on FD002 and FD004 sub-datasets are higher. The reason is that these two sub-datasets contain more operating conditions and fault scenarios but less training samples per case.

The RUL prediction result of the last recorded life cycle on the testing data from the dataset FD001 is shown in Fig.7. Each black dot in Fig.7 is the actual final RUL of each test sample. We sort the 100 testing samples by the actual final RUL values for clearer visualization. The result shows that the RUL estimated by the proposed approach are generally close to the actual RUL values. From Fig. 7, the prediction of RUL has comparatively larger variations for the RUL in the middle range. Because the degradation of the system is slow in this range and then accelerate. The significant change brings difficulty for the deep learning model to learn the sharp change with a limited amount of training samples. Fig. 8 plots the RUL prediction results of the entire life cycles. In this paper, the maximum value of RUL was clipped at 125. This step makes the network to learn from the sequence data more when the engines are close to failing, which has significantly increased the overall prognosis results reported in many literatures [35] [36]. Four degradation test samples from the sub-dataset FD001 are selected to visualize the performance of the developed approach. The result shows that the proposed method can generally make a satisfactory prediction over the entire degradation process. From Fig.8 we can see that the proposed model provides an accurate prediction for the early stage of the degradation as shown in the third subplot and the flat part in the other three subplots. Some errors can be found in the middle periods. Also, in the later stage of the degradation, the RUL prediction has a small variation with the actual value. The prognosis performance in the late phase of the machinery is significant to arrange a proper maintenance action to maintain the system reliability and to reduce the overall cost. The proposed approach achieves superior performance in the RUL prediction of the late phase. Here, some of the RUL predictions are above the actual value and some go under the actual value. One main reason is the inaccuracy of the trained model. The other reason can be the sensing error in the sensory data.

The comparison of RUL prediction results between the proposed method and the existing data-driven approaches is conducted. Fig.9 presents the comparison of the proposed method with other approaches including support vector

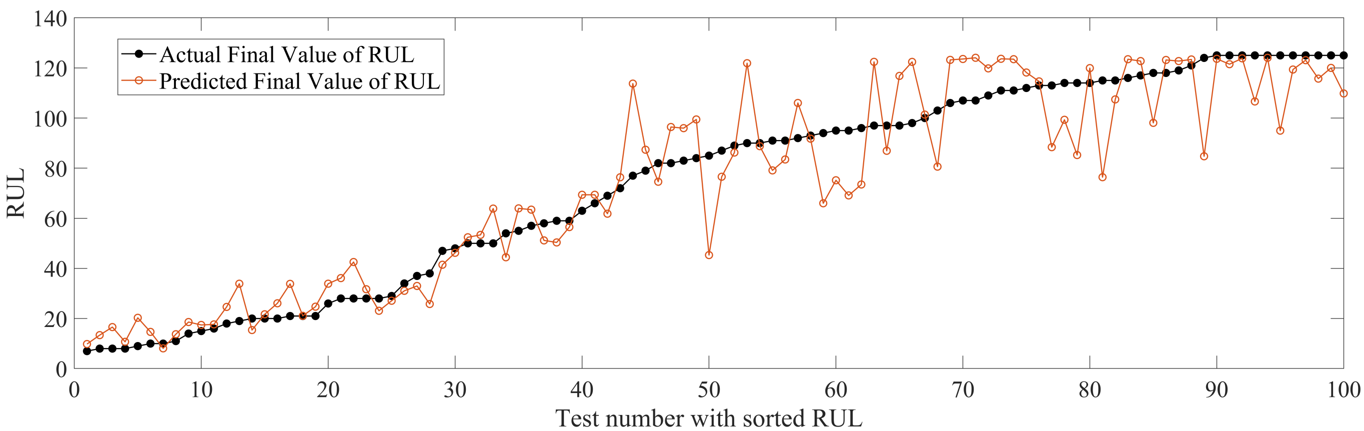

Figure 7: Final RUL value prediction result on the sub-dataset FD001.

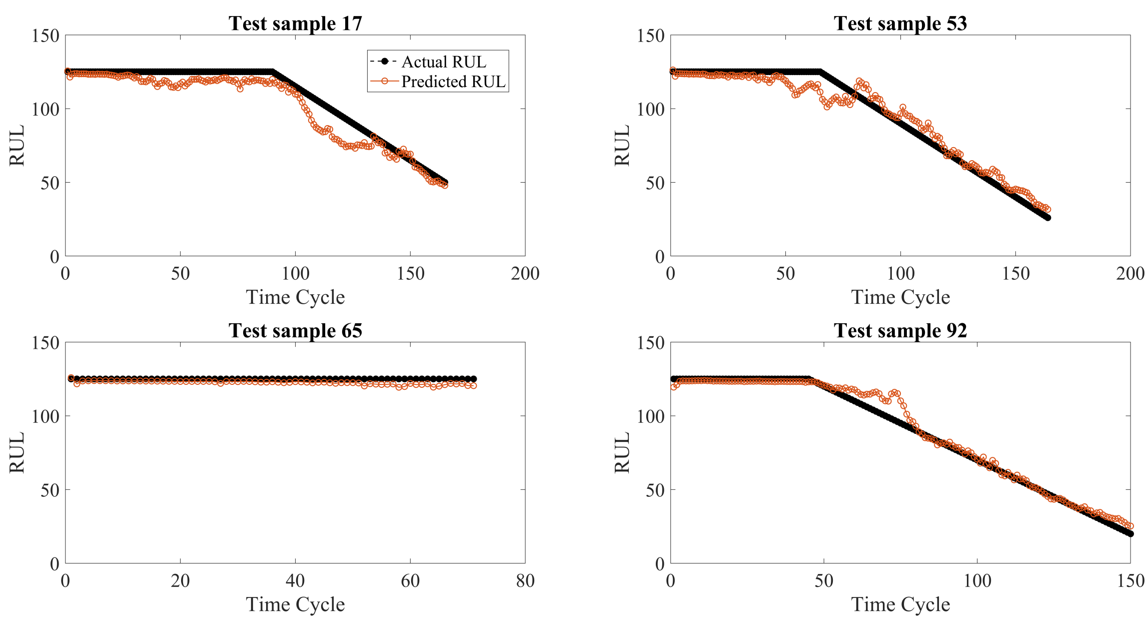

Figure 8: Examples of RUL prediction on the sub-dataset FD001.

machine(SVM) [37], support vector regression [12], random forest [26], convolution neural network (CNN) [38], deep belief network (DBN) [26], and regular LSTM [39]. The proposed approach achieved the best performance. The prediction result of DBN approach is very close to the proposed method. However,

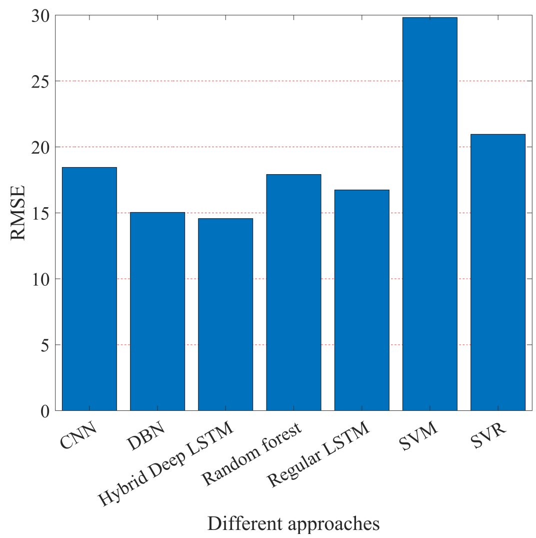

Figure 9: Comparison of RUL predictions on the sub-dataset FD001.

the DBN model did not consider the temporal correlation of the sensory data in prognosis. Besides, the shared parameters of LSTM structure can reduce the requirement of computational power and decrease the chance of overfitting. The developed hybrid deep LSTM approach can be applied directly to the raw sequential data. It overcomes the disadvantages of most of the traditional data-driven prognosis methods where manual feature extraction is needed. The hybrid deep neural network structure can learn the representative features from the raw sensory data directly through the training process. Therefore, the proposed approach can be applied more widely to machines and components where limited prior knowledge is available.

5. Conclusion

In summary, this paper presents a hybrid deep LSTM approach for machinery prognosis. In the hybrid deep neural network, the LSTM layers and the classic neural network layers can capture the information in the sequential sensory data. Besides, the proposed approach achieves sensor fusion by fusing the data from multiple sensors at the data level. The fused data form the input of the model that enhances prognosis performance. The developed method intakes the raw sensory data directly to train the RUL prediction model. Manual feature extraction in most of the traditional data-driven prognosis approaches which requires much prior knowledge and heavy domain expertise can be avoided. To improve the effectiveness of the training process, this paper adopts the dropout technique to prevent overfitting. The decaying learning rate is also used during the training process to increase training efficiency. A comprehensive experimental study is carried out to evaluate the proposed prognosis approach by using the benchmark C-MAPSS dataset. The results show that the prognosis performance is satisfactory using the proposed method. The comparison between the hybrid deep LSTM method and other state-of-art data-driven methods shows the outstanding performance of the developed approach in RUL prediction. Due to the end-to-end learning capability, the developed method can be applied to wide categories of machines and components where limited prior knowledge and domain expertise are available. Future work includes the investigation of the proposed method on the prognosis of machines under more complicated load or working conditions as well as different fault modes. With more sensory data under various operation conditions and fault modes, different types of hybrid deep LSTM models can be evaluated. Also, for the spacial information from the sensory data, the use of the convolution neural network layers into the hybrid structure can be considered to better capture the spacial information.

6. Acknowledgement

This work is partially supported by the Taicang Innovation Leading Project (TC2018DYDS21) and the Deanship of Scientific Research at King Saud University through the research group project number RG-1439-036.

References

[1] E.-H. Aghezzaf, A. Khatab, P. L. Tam, Optimizing production and imperfect preventive maintenance plannings integration in failure-prone manufacturing systems, Reliability Engineering & System Safety 145 (2016) 190 – 198. doi:https://doi.org/10.1016/j.ress.2015.09.017.

[2] C. Su, X. Wang, A two-stage preventive maintenance optimization model incorporating two-dimensional extended warranty, Reliability Engineering & System Safety 155 (2016) 169 – 178. doi:https://doi.org/10.1016/j.ress.2016.07.004.

[3] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, B. Yin, Smart factory of industry 4.0: Key technologies, application case, and challenges, IEEE Access 6 (2018) 6505–6519. doi:10.1109/ACCESS.2017.2783682.

[4] J. Wang, L. Zhang, L. Duan, R. X. Gao, A new paradigm of cloud-based predictive maintenance for intelligent manufacturing, Journal of Intelligent Manufacturing 28 (5) (2017) 1125–1137. doi:10.1007/s10845-015-1066-0.

[5] D. Wang, K. Tsui, Q. Miao, Prognostics and health management: A review of vibration based bearing and gear health indicators, IEEE Access 6 (2018) 665–676. doi:10.1109/ACCESS.2017.2774261.

[6] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to rul prediction, Mechanical Systems and Signal Processing 104 (2018) 799 – 834. doi:https://doi.org/10.1016/j.ymssp.2017.11.016.

[7] S. Yin, X. Li, H. Gao, O. Kaynak, Data-based techniques focused on modern industry: An overview, IEEE Transactions on Industrial Electronics 62 (1) (2015) 657–667. doi:10.1109/TIE.2014.2308133.

[8] M. Xia, T. Li, Y. Zhang, C. W. de Silva, Closed-loop design evolution of engineering system using condition monitoring through internet of things

and cloud computing, Computer Networks 101 (2016) 5 – 18. doi:https://doi.org/10.1016/j.comnet.2015.12.016.

[9] F. Cheng, L. Qu, W. Qiao, Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis (Jan 2018). doi:10.1109/TSTE.2017.2719626.

[10] S. Zhang, Q. He, K. Ouyang, W. Xiong, Multi-bearing weak defect detection for wayside acoustic diagnosis based on a time-varying spatial filtering rearrangement, Mechanical Systems and Signal Processing 100 (2018) 224 – 241. doi:https://doi.org/10.1016/j.ymssp.2017.06.035.

[11] L. R. Rodrigues, J. P. P. Gomes, F. A. S. Ferri, I. P. Medeiros, R. K. H. Galvão, C. L. N. Júnior, Use of phm information and system architecture for optimized aircraft maintenance planning, IEEE Systems Journal 9 (4) (2015) 1197–1207. doi:10.1109/JSYST.2014.2343752.

[12] C. Sbarufatti, M. Corbetta, M. Giglio, F. Cadini, Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks, Journal of Power Sources 344 (2017) 128 – 140. doi:https://doi.org/10.1016/j.jpowsour.2017.01.105.
URL http://www.sciencedirect.com/science/article/pii/S0378775317301155

[13] M. Djeziri, S. Benmoussa, R. Sanchez, Hybrid method for remaining useful life prediction in wind turbine systems, Renewable Energy 116 (2018) 173 – 187. doi:https://doi.org/10.1016/j.renene.2017.05.020.

[14] K. Javed, R. Gouriveau, N. Zerhouni, P. Nectoux, Enabling health monitoring approach based on vibration data for accurate prognostics, IEEE Transactions on Industrial Electronics 62 (1) (2015) 647–656. doi:10.1109/TIE.2014.2327917.

[15] Z. Tian, An artificial neural network method for remaining useful life pre-

diction of equipment subject to condition monitoring, Journal of Intelligent Manufacturing 23 (2) (2012) 227–237. doi:10.1007/s10845-009-0356-9.

[16] P. G. Nieto, E. García-Gonzalo, F. S. Lasheras, F. de Cos Juez, Hybrid pso-svm-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability, Reliability Engineering & System Safety 138 (2015) 219 – 231. doi:https://doi.org/10.1016/j.ress.2015.02.001.

[17] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, N. Zerhouni, Direct remaining useful life estimation based on support vector regression, IEEE Transactions on Industrial Electronics 64 (3) (2017) 2276–2285. doi:10.1109/TIE.2016.2623260.

[18] D. Wu, C. Jennings, J. Terpenny, R. X. Gao, S. Kumara, A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, Journal of Manufacturing Science and Engineering 139 (7) (2017) 071018.

[19] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, X. Wang, Deep learning for pixel-level image fusion: Recent advances and future prospects, Information Fusion 42 (2018) 158 – 173. doi:https://doi.org/10.1016/j.inffus.2017.10.007.

[20] Z. Jianqiang, G. Xiaolin, Z. Xuejun, Deep convolution neural networks for twitter sentiment analysis, IEEE Access 6 (2018) 23253–23260. doi:10.1109/ACCESS.2017.2776930.

[21] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, Z. J. Wang, 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri, IEEE Access 5 (2017) 23626–23636. doi:10.1109/ACCESS.2017.2762703.

[22] M. Xia, T. Li, L. Liu, L. Xu, C. W. de Silva, Intelligent fault diagnosis

approach with unsupervised feature learning by stacked denoising autoencoder, IET Science, Measurement & Technology 11 (2017) 687–695.

[23] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing 115 (2019) 213 – 237. doi:https://doi.org/10.1016/j.ymssp.2018.05.050.

[24] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, Z. Zhu, Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery, IEEE Access 5 (2017) 15066–15079. doi:10.1109/ACCESS.2017.2728010.

[25] M. Xia, T. Li, L. Xu, L. Liu, C. W. de Silva, Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks, IEEE/ASME Transactions on Mechatronics 23 (1) (2018) 101–110. doi:10.1109/TMECH.2017.2728371.

[26] C. Zhang, P. Lim, A. K. Qin, K. C. Tan, Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics, IEEE Transactions on Neural Networks and Learning Systems 28 (10) (2017) 2306–2318. doi:10.1109/TNNLS.2016.2582798.

[27] M. Xia, T. Li, L. Liu, L. Xu, S. Gao, C. W. de Silva, Remaining useful life prediction of rotating machinery using hierarchical deep neural network, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 2778–2783. doi:10.1109/SMC.2017.8123047.

[28] M. Kolbaek, D. Yu, Z.-H. Tan, J. Jensen, M. Kolbaek, D. Yu, Z.-H. Tan, J. Jensen, Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks, IEEE/ACM Trans. Audio, Speech and Lang. Proc. 25 (10) (2017) 1901–1913. doi:10.1109/TASLP.2017.2726762.

[29] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocom-

puting 240 (2017) 98 – 109. doi:https://doi.org/10.1016/j.neucom.2017.02.045.

[30] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional lstm and other neural network architectures, Neural Networks 18 (5) (2005) 602 – 610. doi:https://doi.org/10.1016/j.neunet.2005.06.042.

[31] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Exploring recurrent neural networks for on-line handwritten signature biometrics, IEEE Access 6 (5128-5138) (2018) 1–7.

[32] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, J. Schmidhuber, Lstm: A search space odyssey, IEEE Transactions on Neural Networks and Learning Systems 28 (10) (2017) 2222–2232. doi:10.1109/TNNLS.2016.2582924.

[33] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747 (2016).

[34] A. Saxena, K. Goebel, D. Simon, N. Eklund, Damage propagation modeling for aircraft engine run-to-failure simulation, in: 2008 International Conference on Prognostics and Health Management, 2008, pp. 1–9. doi:10.1109/PHM.2008.4711414.

[35] S. K. Singh, S. Kumar, J. P. Dwivedi, A novel soft computing method for engine rul prediction, Multimedia Tools and Applications (Sep 2017). doi:10.1007/s11042-017-5204-x.

[36] J. Zhang, P. Wang, R. Yan, R. X. Gao, Long short-term memory for machine remaining life prediction, Journal of Manufacturing Systems 48 (2018) 78 – 86, special Issue on Smart Manufacturing. doi:https://doi.org/10.1016/j.jmsy.2018.05.011.
URL http://www.sciencedirect.com/science/article/pii/S0278612518300803

[37] C. Louen, S. X. Ding, C. Kandler, A new framework for remaining useful life estimation using support vector machine classifier, in: 2013 Conference on Control and Fault-Tolerant Systems (SysTol), 2013, pp. 228–233. doi:10.1109/SysTol.2013.6693833.

[38] G. Sateesh Babu, P. Zhao, X.-L. Li, Deep convolutional neural network based regression approach for estimation of remaining useful life, in: Database Systems for Advanced Applications, Springer International Publishing, 2016, pp. 214–228.

[39] C. Hsu, J. Jiang, Remaining useful life estimation using long short-term memory deep learning, in: 2018 IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 58–61. doi:10.1109/ICASI.2018.8394326.