A Novel Similarity-based Method for Remaining Useful Life Prediction Using Kernel Two Sample Test

Xiaodong Jia[1], Haoshu Cai[2], Yuanming Hsu[3], Wenzhe Li[4], Jianshe Feng[5] and Jay Lee[6]

[1,2,3,4,5,6,] *NSF I/UCR Center for Intelligent Maintenance Systems, Department of Mechanical Engineering, University of Cincinnati, PO Box 210072, Cincinnati, Ohio 45221-0072, USA*

jiaxg@mail.uc.edu
caihu@mail.uc.edu
hsuyg@mail.uc.edu
li2w6@mail.uc.edu
fengje@mail.uc.edu
lj2@ucmail.uc.edu

ABSTRACT

When abundant historical Run-to-Failure (R2F) data is available, the similarity-based method is one of the favored options for Remaining Useful Life (RUL) prediction due to its simplicity and satisfactory accuracy. In this study, a novel similarity-based methodology for RUL prediction is proposed. The proposed method has two important processing steps: similarity matching and Weibull fitting. The similarity matching screens the historical records by a similarity testing called Kernel Two Sample Test (KTST), and only those records that pass KTST are adopted as references for RUL prediction. For the selected similar records, the RUL is predicted as the remaining time to failure. The Weibull fitting fuses the multiple RUL predictions given by similar historical records. The PDF of RUL is estimated as the fitted Weibull distribution. To demonstrate the effectiveness and superiority of the proposed method, the famous C-MAPSS (Modular Aero-Propulsion System Simulations) data about aero-engine degradation is adopted for model validation. The results demonstrate improved prediction accuracy comparing with other similarity-based approaches and the state-of-the-art deep learning predictors.

1. INTRODUCTION

Remaining Useful Life (RUL) prediction is an integral part of Prognostics and Health Management (PHM) that aims to predict the remaining time to failure of the machine based on the condition monitoring data. Normally, the RUL prediction serves as important inputs to the subsequent maintenance planning and optimization. In practice, the RUL of the machine is often predicted as a Probability Density Function (PDF). The mean value of the PDF is the predicted RUL and the confidence interval at specified significance level is used to describe the prediction uncertainty. Existing approaches for RUL prediction include RBM, RCM, SPM, SSM and SBM, as in Table I.

Among these prediction methods, the similarity-based method is one of most favored and intuitive methods due to its simplicity for implement. However, this method is still less discussed and investigated in the literature than other approaches. Based on the literature review, we identified several reasons that may possibly limit the popularity and practicality of SBMs. 1) most SBMs, such as Trajectory Based Prediction Method (TSBP) and RULCLIPPER (Remaining Useful Life estimation based on impreCise heaLth Indicator modeled by Planar Polygons and similarity-basEd Reasoning), are Health Indicator (HI) based and they require HI estimation as a pre-processing step. However, the uncertainty of HI estimation is hard to quantify, and it may introduce additional disturbances to the final RUL prediction. 2) There are limited similarity criteria can evaluate the similarity between two multivariate temporal sequences. Although Mahalanobis distance may be potentially useful, it requires matrix inversion and it is less robust. More importantly, it is difficult to design a threshold to decide whether the two data sets are similar; 3) Based on the current literature, it is not clear what is a better way than weighted averaging to fuse the multiple RUL predictions based on historical samples. 4) The prediction uncertainty of most SBMs are not discussed.

To tackle these challenges, this study proposed a novel similarity-based method for RUL prediction based on Kernel

Xiaodng Jia *et al.* This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Table I A review of RUL prediction methods

	Candidate Algorithms	Model	Pros	Cons
Regression Based Method (RBM)	Neural networks (NN) Ensemble regressors Covariate Based Hazard Models (CBHM)	***M1*** ***M2*** ***M3***	1) Simple and efficient; 2) Good accuracy when deep learning models are employed.	1) The prediction results are inconsistent over different runs due to the random initialization; 2) Lack of uncertainty description 3) The model assumes the underlying degradation is linear over time or operation cycles.
Random Coefficient Method (RCM)	Exponential model	***M2***	1) Sound statistical interpretation of prediction uncertainty;	1) Require HI to be directly observable; 2) Not applicable to multivariate temporal data; 3) Need prior assumptions about the degradation trend;
Stochastic Process Method (SPM)	Weiner processes Gamma processes	***M2***	1) Sound statistical interpretation of prediction uncertainty;	1) Require HI to be directly observable; 2) Not applicable to multivariate temporal data; 3) Need prior assumptions about the degradation trend;
State Space Method (SSM)	Particle Filters (PF) Kalman Filters (KF) Hidden Markov Model (HMM) Monte Carlo Filters	***M3***	1) Does not require HI to be observable; 2) Can be applied to multivariate series; 3) Good interpretation of uncertainty; 4) Robust to noisy HI;	1) Need assumptions about degradation trend; 2) Need to estimate HI from the observable data; 3) The uncertainty of HI estimation is difficult to quantify;
Similarity Based Method (SBM)	Trajectory Similarity Based Prediction (TSBP) RULCLIPPER Match matrix	***M1*** ***M3***	1) Simple and efficient; 2) Good accuracy when abundant historical data is available	1) Lack of uncertainty descriptions; 2) Need abundant data to make predictions; 3) Less robust in some cases;

Two Sample Test (KTST). Comparing with the existing discussions about SBM, the proposed method holds several advantages. 1) The proposed method is not HI based, and the similarity between two multivariate temporal sequences is directly evaluated by Maximum Mean Discrepancy (MMD); 2) The KTST (Gretton, Borgwardt, Rasch, Schölkopf, & Smola, 2012) is adopted to obtain an upper bound for the MMD to decide which R2F profile in the historical database is similar to the current data. Only similar data samples are employed as references to make predictions. 3) Weibull analysis is used to fuse multiple RUL predictions given by the historical data records and the PDF of RUL is obtained as the fitted Weibull distribution.

The remainder of this paper is organized as follows. Section 2 give a brief review of RUL prediction methods and an introduction about MMD and KTST. Section 3 details the proposed method and necessary pre-processing steps. Section 4 validates the proposed method based on the famous CMAPSS data for aero-engine RUL prediction. The conclusion remarked are presented in Section 5.

2. LITERATURE REVIEW AND TECHNICAL BACKGROUND

2.1. Review of RUL Prediction

In the literature, the available mathematical models for RUL prediction can be summarized into three types as follows:

M1: Direct RUL prediction model: the model predicts the RUL directly based on the machine operation data;

M2: Direct HI-based prediction model: the model predicts the RUL by extrapolating a measurable HI to a targeted failure threshold, as shown in Fig. 1. The model normally requires a known failure threshold to make RUL predictions.

M3: Indirect HI-based prediction model: In this prediction model, the HI of the system is not observable. Therefore, the model estimates HI from the machine operation data first and then predicts RUL using the indirect HI based prediction models in ***M2***.

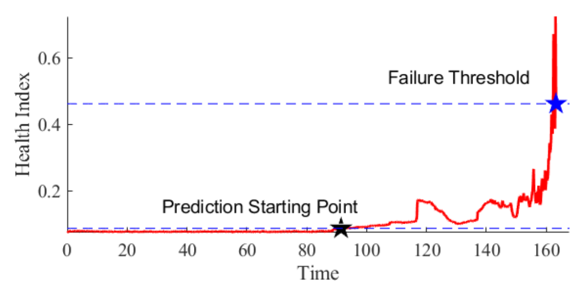

Fig. 1 HI based (or sensor based) RUL prediction

We note that the biggest difference between ***M2*** and ***M3*** is whether the HI is directly measurable. For most engineering systems, the health of the system is not directly observable. Therefore, the HI estimation in ***M3*** serves as a virtual health measurement system to infer the underlying degradation pattern of the machine.

The above-mentioned prediction models can be implemented by using different prediction methods, as tabulated in Table I. In Table I, the simplest method for RUL prediction is the RBM which establishes a direct mapping from the machine data to the RUL using off-the-shelf regression techniques, such as SVR (Support Vector Regression)(Khelif et al., 2016), NN (Shao & Nezu, 2000), etc. When training the model, the RUL is used as the target variable of the regressor. However, this implicitly assumes the machine degradation is linear, since the RUL decreases linearly over time or operation cycles. Therefore, the RBM is recommended for initial trials to establish baseline prediction accuracies for future improvements. The prediction results might be seriously biased when the system degradation is highly nonlinear.

The RBM can be also adopted for sensor-based RUL prediction. Typical examples involve using ARMA (Auto-Regressive Moving Average) (Liao & Köttig, 2014), SVR to extrapolate the sensor readings into future horizons. In this case, the regression techniques are used for time series prediction. In addition, the RBM can be employed to estimate HI based on the machine data. Relevant examples can be read in (Jain & Lad, 2016), which employ regression techniques to estimate the degradation of milling cutters. In this study, the regression model serves as a virtual health metrology device to replace the expensive camera system for milling cutter monitoring.

The RCM are wide adopted to obtain the Probability Density Function (PDF) of RUL. An early study by Lu and Meeker (C. J. Lu & Meeker, 1993) demonstrates a statistical method to estimate the RUL distribution for a broad class of degradation model. The noise term in (C. J. Lu & Meeker, 1993) is additive noise and it follows a zero-mean normal distribution. Following the work in (C. J. Lu & Meeker, 1993), several variants and applications of the random coefficient model can be found in (J. C. Lu, Park, & Yang, 1997; Tseng, Hamada, & Chiao, 1995; Upadhyaya, Naghedolfeizi, & Raychaudhuri, 1994; Yang & Jeang, 1994). Gebraeel *et al.* (Gebraeel, Lawley, Li, & Ryan, 2005) investigates the two-parameter exponential models with multiplicative random error terms and with multiplicative Brown motion error. In this paper, closed-from solutions for the exponential models are obtained using a Bayesian approaches. Later investigations in (Chakraborty, Gebraeel, Lawley, & Wan, 2009; A. Elwany & Gebraeel, 2009; A. H. Elwany & Gebraeel, 2008; Kaiser & Gebraeel, 2009) that are given by Gebraeel and his co-authors presented several extensions and improvements based on the exponential model in (Gebraeel et al., 2005). A more recent study by X. Si *et al.* (Si, Wang, Chen, Hu, & Zhou, 2013) proposes a path-dependent model for adaptive RUL estimation by considering the online updating of degradation model from the most-updated observations.

The SPM describes the degradation process as a stochastic process. The model structure of SPM is quite like RCM. The difference is that the noise term is modeled as a stochastic process in SPM (Si et al., 2013). In the literature, Wiener processes and gamma processes are widely investigated for RUL predictions and HI predictions. A recent review that is given by Z.Zhang *et al.*(Z. Zhang, Si, Hu, & Lei, 2018) presents an excellent review of using wiener processes for degradation data analysis and RUL estimations.

SSM is another family of prediction method for RUL prediction, which is essentially HI based. The measurement equation in SSM maps the machine data or noisy HI observations to a latent variable. The state function updates the state estimates first and then predicts the future propagation of failure. To give a few examples, D. An *et al.* (An, Choi, & Kim, 2013) presented a tutorial for particle filter-based prognostics algorithms and applies it to the Li-on battery degradation prediction. J. Sun *et al.* (Sun, Zuo, Wang, & Pecht, 2012, 2014) uses Kalman filters to estimate the RUL distribution of aero-engines based on the public data given by CMAPSS.

The similarity-based method for RUL prediction is also studied when abundant historical data is available. The TSBP was introduced in the PHM data competition 2008 (Wang, 2010) and won this data competition. In this analysis, the Logistical regression is used to estimate the HI of the engine, the similarity was evaluated based on the Euclidean distance between current and historical HIs. Another recent study in proposes to use Multi-Task Gaussian Process (MTGP) to achieve the reference-based prediction of the SoH of Li-on batteries. Their method demonstrates significant improvements in the prediction accuracies.

2.2. Kernel Two Sample Test

MMD is a similarity statistic that evaluates the discrepancy of two distributions (Gretton, Borgwardt, Rasch, Schölkopf, & Smola, 2007a, 2007b). The biased estimate of MMD can be described as:

$$\begin{aligned}\mathrm{MMD}_b^2 &= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}k(x_i,x_j)+\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}k(y_i,y_j)\\ &-\frac{2}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}k(x_i,y_i)\end{aligned} \tag{1}$$

Where x_i is the i-th sample from $\mathbf{X}=\{x_1,\dots,x_m\}$, y_i is the i-th sample from $\mathbf{Y}=\{y_1,\dots,y_m\}$. Elements in **X** and **Y** are $i.i.d$ (identically independent distributed) random samples are drawn from distribution p and q respectively. The subscript b in Eq. (1) denotes the biased estimate of MMD. The kernel function $k(\cdot,\cdot)$ in Eq. (1) is a Radial Basis Kernel (RBF) kernel with length scale equals to 1 for all the studies.

The statistical meaning of MMD is the distance between mean embedings of two distributions in kernel space. A large value of MMD means the two distributions are dissimilar, wheares a small value of MMD means the two distributributions are similar.

Kernel two sample test is a hypothesis test that is based on MMD. The null hypothesis of the kernel two sample test states that **X** and **Y** are sampled from different distributions. By rejecting the null hypothesis, **X** and **Y** follow the same distribution. Under the null hypothesis, these test statistic (MMD) are expected to be close to 0, and the values of these statistics should be smaller than the test bounds given by the kernel two sample test (Gretton et al., 2012). The theoretical explanation of the kernel two sample test is explained in (Gretton et al., 2012). The basic idea of KTST is to model the MMD value as a gamma distribution, as shown in Fig. 2. The test upper bound is obtained by calculating the quantile at $1-\alpha$, where α is the significance level that is specified by users. In default setting, $\alpha = 0.05$.

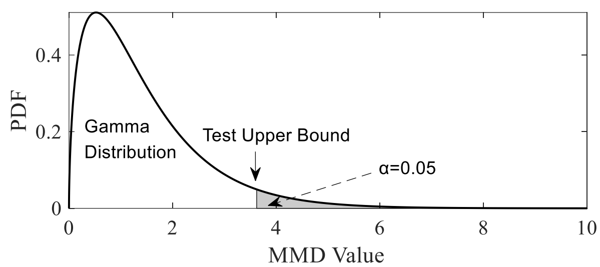

Fig. 2 Illustration of tuning parameter $\boldsymbol{\alpha}$

3. METHODOLOGY DEVELOPMENT

By reviewing the available methods and analytics for RUL prediction, this paper contributes a novel similarity-based method using KTST. Comparing the existing approaches, we highlight the advantages of presented method as follows. 1) the proposed method avoids the ambiguous HI estimation step; 2) The MMD in this study directly evaluates the similarity between the sensor readings; 3) The KTST screens the dissimilar historical data. Therefore, the RUL prediction in the proposed method is only based on the ones that can pass the similarity test. 4) The uncertainty of the prediction is described by a Weibull distribution, which is widely employed for life data analysis in reliability engineering.

3.1. An Overview of the Proposed Methodology

An overview of the proposed method is shown in Fig. **3**, which has three key steps. The similarity matching finds the similar historical degradation profiles in the database using KTST. Only the historical data that passes the KTST are regarded as similar and are employed for RUL prediction. The dissimilar data are excluded from the RUL prediction. We also note that the historical degradation profiles in the database must be Run-to-Failure data that cover the whole life cycle of the machine. In the following discussion, the historical database is denoted as $\boldsymbol{D} = \{d_k\}_{k=1,\dots,N}$, where d_k is the k-th R2F profile and N is the total number of historical profiles. After identifying the similar peers in history, a set of RUL predictions are subsequently obtained. In the following step, these predictions based on historical data are fitted into a Weibull distribution to describe the PDF of the RUL.

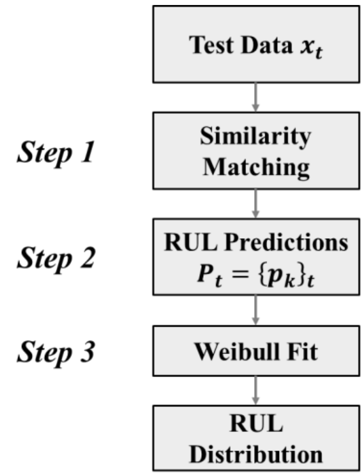

Fig. 3 Flowchart of the proposed methodology

Similarity matching is the key step in the proposed method to identify the similar profiles from historical database. A detailed illustration of the similarity matching is in Fig. 4. In this flowchart, the test data is first truncated using a time window and only the most recent observations are adopted for similarity matching. The windowed test data is denoted as x_t and the length of the time window is B. Next, a sliding window with length B is applied to historical data d_k. The data in this sliding window is compared with the test data x_t to find the best match. The best match is obtained by minimizing MMD value along the time axis and the time at the best match is denoted as t^k_{match} , as in Fig. 4. Since d_k is a R2F failure profile, the time at the End of Life (EoL) is t^k_{EoL} and the predicted RUL based on d_k at time t is written as:

$$P_{k,t} = t^k_{EoL} - t^k_{match} \tag{2}$$

After finding the best matching, the windowed data from d_k and x_t needs to pass the KTST to be adopted for final RUL prediction. If the KTST is not rejected, the test data x_t and the historical data d_k is regarded as dissimilar and the RUL prediction P_k is discarded.

3.2. Feature Selection

Feature selection is an important pre-processing step for the proposed method, which may have large impact on the prediction performance. In the present case, useful features for RUL prediction are selected by checking the monotonicity and consistency of individual feature based on a population of R2F data profiles.

The monotonicy test we based on is Mann-Kendall (MK) test (Mann, 1945), which is designed to assess if there is a monotonic trend of the series over time. Under null hypothesis, the data is not monotonic. Application of MK test

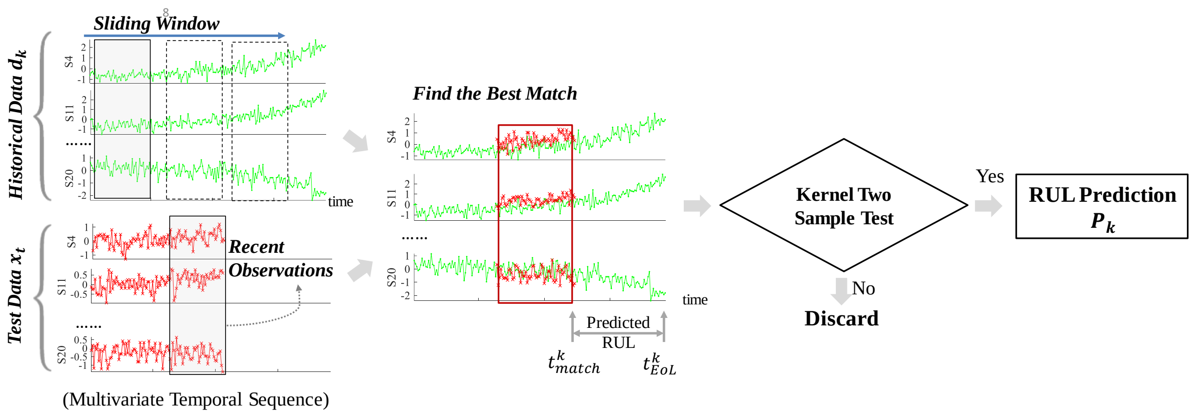

Fig. 4 Process of similarity matching

in boring tools and ballscrew degradation problem can be read from our previous work in (Jia, Zhao, Di, Yang, & Lee, 2018) and (P. Li et al., 2018). Given an individual feature, the monotonicity of the feature is quantified by the pass rate of MK test in a population of R2F cycles. If the pass rate is larger than 95%, it is regarded as monotonic. Otherwise, it will be excluded from prediction. We note that the feature monotonicity is just a basic requirement. To be useful for RUL prediction, the feature consistency over different life cycles needs to be further evaluated.

The feature consistency is quantified by the uncertainty score that is written as follows:

$$s = \exp\left(-\frac{std(h_{EoL})}{mean(|h_{EoL} - h_0|)}\right) \tag{3}$$

where h_{EoL} represents all the final health values at EoL and h_0 means the HI at the beginning of the life cycle. The index s quantitatively measures the feature consistency over different life cycles and this value ranges from 0 to 1. When s is close to 1, it means the consistency of the feature is good. To the author's experience, $s \geq 0.75$ means the feature is useful for RUL prediction.

3.3. Parameter Tuning

The length of time window B and the significance level α for KTST are the two tuning parameters in the proposed method. It is found in practice that a larger time window tends to give more consistent prediction results, since more data points are considered in the similarity matching. However, if the time window B is too large, the similarity criterion might be too strict and only very limited historical data can pass the similarity testing. On the contrary, if B is too small, more historical data will pass the similarity testing, which may enlarge the uncertainty of final prediction. Like other methods, the selection of time window length B is quite intuitive and is data dependent. For the CMAPSS data, the recommended window length is $\mathrm{B} = 100$, and this value is suggested to be larger than 60.

As for the significance level, $\alpha = 0.05$ is recommended for most engineering practices. As in Fig. 2, a larger value of α means a strict upper bound for the KTST. In the following discussion, $\alpha = 0.05$ is adopted.

We would like to note that the parameter tuning for similarity-based RUL prediction is meaningless. For most machine learning applications, the parameters are tuned by cross-validation, such as K-fold, Leave-One-Out (LOO), etc., on the training set. However, it is not clear how to implement cross-validation in the present case, since the data records in the R2F database are multivariate temporal sequences rather than vector-based data samples. If the parameters are tuned based on the testing set, the results might be seriously biased. Therefore, we test our method under default setting without any parameter tuning. We also note that the cross-validation on the training set is applicable when the regression-based methods are adopted for direct RUL prediction.

4. RESULTS AND DISCUSSIONS

The effectiveness of the proposed method is demonstrated based on the CMAPSS dataset for aero-engine RUL prediction. The given data in the training set are multiple multivariate temporal sequences that describe the life time degradation of the areo-engines. Each multivariate series contains 21 sensor readings. The data records in the testing set are partial degraded data and RUL for each unit is unknown.

Table II Data description

Unit	Failure Mode	Conditions	N_{train}	N_{test}
FD001	HPC (High Pressure Compressor)	1	100	100
FD002	HPC	6	260	259
FD003	HPC, Fan	1	100	100
FD004	HPC, Fan	6	249	248
Total	--	--	709	707

Table III Benchmarking of prediction accuracies

		Proposed Method	DCNN	LSTM	RULCLIPPER	**Proposed Method**	MODBNE	DBN	RF	GB	SVM	LASSO
FD001	RMSE	**16.43**	12.61	13.52	13.27	**19.03**	15.04	15.21	17.91	15.67	40.72	19.74
	Score	**369**	273	431	216	**622**	334	417	479	474	7703	653
FD002	RMSE	**23.36**	22.36	24.42	22.89	**36.55**	25.05	27.12	29.59	29.07	52.99	37.13
	Score	**2671**	14459	10412	2796	**74870**	5585	9031	70465	87280	316483	276923
FD003	RMSE	**17.43**	12.64	13.54	16.00	**26.62**	12.51	14.71	20.27	16.84	46.32	21.38
	Score	**1129**	284	347	317	**12697**	421	442	711.13	576	22541	1058
FD004	RMSE	**23.36**	23.31	24.21	24.33	**30.86**	28.66	29.88	31.12	29.01	59.96	40.70
	Score	**2670**	12466	14322	3132	**31277**	6557	7954	46567	17817	141122	125297
R_{early}		**125**	125	125	135	**Not Applied**	Not Applied	Not Applied	Not Applied	Not Applied	Not Applied	Not Applied

A summary of the CMAPSS is presented in Table II. There are four different units running under different working conditions. The data set FD001 and FD002 has only one working condition but less historical records than another two units. In comparison, the data set FD002 and FD004 have 6 different working conditions.

In the pre-process steps, the data is firstly normalized to zero-mean and unitary variance. As for FD002 and FD004, the data under different working conditions are normalized separately. After data normalization, sensor selection is implemented by evaluating the monotonicity and uncertainty of individual features. Fig. 5 (a) shows the pass rate of MK test for individual sensor channel and Fig. 5 (b) shows the uncertainty score that is calculated based on Eq.(**3**). It is found that S2, S3, S4, S7, S8, S9, S11, S12, S13, S15, S17, S21, S21 are monotonic features. However, S8, S9, S13, S14 are not recommended for RUL prediction, since these sensor recordings have large uncertainty. To confirm this finding, we visualize the S9 and S11 in Fig. 6. One can clearly see that S11 have good consistency over different life cycles and the values of h_{EoL} demonstrate a spiky distribution centered around 2.1 in the y-axis. In comparison, S9 has large uncertainty which makes it less useful for RUL prediction.

(a)

(b)

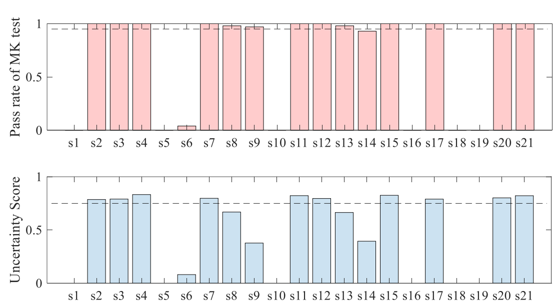

Fig. 5 Sensor selection for train set FD001 in CMAPSS data

Fig. 7 gives two examples of typical prediction trajectories for training data. These two units are randomly selected from the training set of FD004. The findings from Fig. 7 can be summarized as follows. 1) The proposed method suffers large uncertainty at an early stage of degradation. However, the prediction becomes more and more accurate as the engine unit approaches the EoL. 2) The proposed method can describe the PDF of RUL well, especially at the late stage of degradation. Based on the results, there is a high chance that the ground truth falls within the 25%~75% confidence interval and the 5% confident limit can be effectively employed to guide the maintenance activities. 3) The PDFs of RUL in Fig. 7 can be plugged into an optimization model to minimize operation risks and to optimize the costs.

(a)

(b)

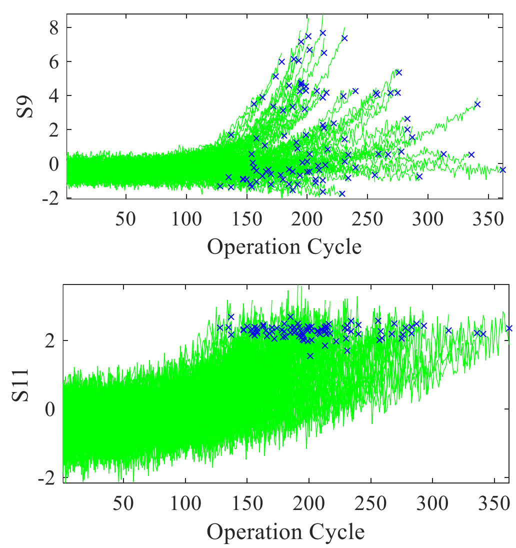

Fig. 6 visualization of the features for train set FD001. Green line are the sensor recordings and blue cross shows the value of $\boldsymbol{h_{EoL}}$.

Table III benchmarks the proposed method with existing approaches in detail. R_{early} in Table III is a RUL threshold which has a noticeable effect on the prediction accuracy. As in Fig. 7, the prediction results limits the $R_{early} = 125$. This is because the RUL prediction is only meaningful after the incipient signature of degradation is detected, which widely referred as starting point of RUL prediction, as in Fig. 1. In the literature, R_{early} for CMAPSS data is normally set to 125(Li, Ding, & Sun, 2018), 135(Ramasso, 2014) or not applied(C. Zhang, Lim, Qin, & Tan, 2017). In Table III, DCNN, LSTM, DBN, MODBNE are NN based regressors. RF and GB are ensemble regressors. SVM and LASSO are sparse regularized regression techniques. RULCLIPPER is a similarity-based RUL prediction method. The prediction accuracy for DCNN (Deep Convoluted Neural Networks) and LSTM (Long Short Term Memory neural network) are reported in (X. Li et al., 2018), the accuracy for RULCLIPPER is reported in (Ramasso, 2014) and the accuracies for MODBNE (Multi-Objective Deep Belief Networks Ensemble), DBN (Deep Belief Networks), RF(Random Forest), GB(Gradient Boosting), SVM(Support Vector Machine), Lasso are reported in (C. Zhang et al., 2017). In this investigation, the studies that belong to the family of RCM, SPM and SSM are not benchmarked, since these methods regard the current degradation path as independent from the historical records.

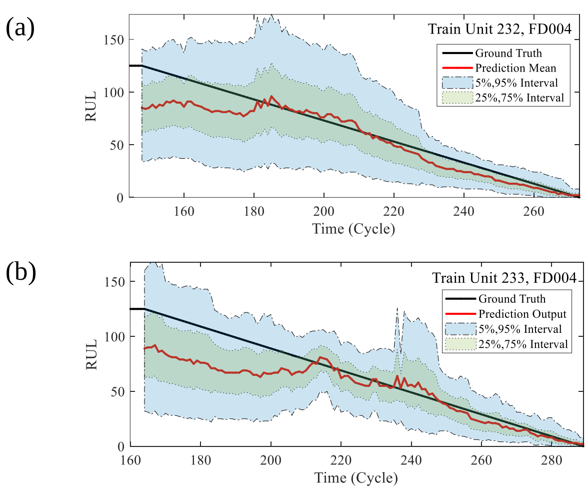

Fig. 7 Examples of life time RUL prediction

When $R_{early} = 125$ is applied, the findings from the benchmarking can be summarized as follows. 1) Comparing with DCNN and LSTM, the proposed method indicates improved score on FD002 and FD004. This is because more historical records are available for RUL prediction. In comparison, the prediction results on FD001 and FD004 are less competitive than DCNN and LSTM, since less historical samples are available than FD002 and FD004. Especially, the prediction results on the Fan degradation in FD003 is not satisfactory, since only limited historical samples have FAN degradation in the training set of FD003. 2) Unlike the DCNN, LSTM and other NN based methods, the proposed method does require random initialization and thus the prediction result is deterministic as long as the training set if fixed. 3) Comparing with RULCLIPPER, the proposed method is much simpler for implementation and there are fewer tuning parameters. Moreover, the tabulated prediction accuracy in Table III is tuned to the best performance based on the testing set, which we believe is questionable in applications.

When R_{early} is not applied, the proposed method is more accurate than off-the-shelf regression predictors, such as RF, GB, SVM, LASSO. However, it is less accurate than the deep neural network such as DBN (Deep Belief Networks). This is essentially because the proposed method indicates large prediction error at an early stage of degradation, which is also visualized in Fig. 7. This is explainable since the similarity-based method makes predictions based on the similarity between the degraded trends. If the machine has no sign of degradation, then this approach amounts to the traditional Weibull analysis, which may have large prediction error on specific unit. Comparing with the off-the-shelf regressors, the proposed method performs better on FD002 and FD004. This is mainly because more historical records are available for prediction. Therefore, we can conclude that the proposed method can become more accurate if more historical records are adopted as reference for prediction. However, as the size of historical database grows, the searching complexity increases quadratically due to the computation complexity of MMD. Although the linear complexity MMD is available, we found the prediction results are not satisfactory.

5. CONCLUSION

In summary, this paper proposes a novel similarity-based approaches for RUL prediction. The proposed method has two important processing steps: similarity matching and Weibull fitting. The similarity matching screens the historical records by similarity and only those records that pass KTST are adopted as references for RUL prediction. The Weibull fitting fuses the multiple RUL predictions given by similar historical records and estimates the PDF of RUL as the fitted Weibull distribution. The Effectiveness of the proposed is demonstrated based on the CMAPSS data for aero-engine RUL prediction. By benchmarking with existing approaches, the pros and cons of the proposed method are summarized as follows:

Pros:

1) Satisfactory prediction accuracy comparing with existing approaches;
2) Less tuning parameters and simplified pre-treatment comparing with other similarity-based approaches, such as RULCLIPPER, TSBP;
3) Good interpretation of prediction uncertainties comparing with the regression-based methods and the similarity-based methods;
4) Unlike the NN based methods which requires random initialization, the prediction results given by proposed method is consistent over different runs;

5) Comparing with the SSM and random-coefficient methods, the proposed method does not require HI to be directly observed or to be estimated from multivariate temporal data.
6) Comparing with the HI based methods, the proposed method does not have ambiguity of setting failure threshold.

Cons:

1) Less accurate when prediction is made at an early stage of degradation.
2) The algorithm requires large amount data (or data with enough diversity) to make accurate predictions;
3) Searching complexity increase quadratically as the size of the historical database increases.

In the future direction for improvement, the similarity-based pre-diagnosis of engine degradation mode will be investigated and integrated to proposed method, which is expected to improve the prediction accuracy significantly.

REFERENCES

An, D., Choi, J. H., & Kim, N. H. (2013). Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. *Reliability Engineering & System Safety, 115*, 161-169. doi:10.1016/j.ress.2013.02.019

Chakraborty, S., Gebraeel, N., Lawley, M., & Wan, H. (2009). Residual-life estimation for components with non-symmetric priors. *IiE Transactions, 41*(4), 372-387. doi:Pii 908239981
10.1080/07408170802369409

Elwany, A., & Gebraeel, N. (2009). Real-time estimation of mean remaining life using sensor-based degradation models. *Journal of manufacturing science and engineering, 131*(5), 051005.

Elwany, A. H., & Gebraeel, N. Z. (2008). Sensor-driven prognostic models for equipment replacement and spare parts inventory. *IiE Transactions, 40*(7), 629-639. doi:10.1080/07408170701730818

Gebraeel, N. Z., Lawley, M. A., Li, R., & Ryan, J. K. (2005). Residual-life distributions from component degradation signals: A Bayesian approach. *IiE Transactions, 37*(6), 543-557.

Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B., & Smola, A. J. (2007a). *A kernel approach to comparing distributions.* Paper presented at the PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE.

Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B., & Smola, A. J. (2007b). A kernel method for the two-sample-problem. *Advances in neural information processing systems, 19*, 513.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research, 13*(Mar), 723-773.

Jain, A. K., & Lad, B. K. (2016). Data driven models for prognostics of high speed milling cutters. *International Journal of Performability Engineering, 12*(1), 03-12.

Jia, X. D., Zhao, M., Di, Y., Yang, Q. B., & Lee, J. (2018). Assessment of Data Suitability for Machine Prognosis Using Maximum Mean Discrepancy. *IEEE Transactions on Industrial Electronics, 65*(7), 5872-5881. doi:10.1109/Tie.2017.2777383

Kaiser, K. A., & Gebraeel, N. Z. (2009). Predictive Maintenance Management Using Sensor-Based Degradation Models. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 39*(4), 840-849. doi:10.1109/Tsmca.2009.2016429

Khelif, R., Chebel-Morello, B., Malinowski, S., Laajili, E., Fnaiech, F., & Zerhouni, N. (2016). Direct Remaining Useful Life Estimation Based on Support Vector Regression. *IEEE Transactions on Industrial Electronics*, 1-1. doi:10.1109/tie.2016.2623260

Li, P., Jia, X. D., Feng, J. S., Davari, H., Qiao, G., Hwang, Y., & Lee, J. (2018). Prognosability study of ball screw degradation using systematic methodology. *Mechanical Systems and Signal Processing, 109*, 45-57. doi:10.1016/j.ymssp.2018.02.046

Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety, 172*, 1-11. doi:10.1016/j.ress.2017.11.021

Liao, L., & Köttig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability, 63*(1), 191-207.

Lu, C. J., & Meeker, W. O. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics, 35*(2), 161-174.

Lu, J. C., Park, J., & Yang, Q. (1997). Statistical inference of a time-to-failure distribution derived from linear degradation data. *Technometrics, 39*(4), 391-400. doi:Doi 10.2307/1271503

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, 245-259.

Ramasso, E. (2014). Investigating computational geometry for failure prognostics. *International Journal of Prognostics and Health Management, 5*(1), 005.

Shao, Y., & Nezu, K. (2000). Prognosis of remaining bearing life using neural networks. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 214*(3), 217-230.

Si, X.-S., Wang, W., Chen, M.-Y., Hu, C.-H., & Zhou, D.-H. (2013). A degradation path-dependent approach for remaining useful life estimation with an exact and closed-form solution. *European Journal of Operational Research, 226*(1), 53-66.

Sun, J., Zuo, H., Wang, W., & Pecht, M. G. (2012). Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing, 28*, 585-596.

Sun, J., Zuo, H., Wang, W., & Pecht, M. G. (2014). Prognostics uncertainty reduction by fusing on-line monitoring data based on a state-space-based degradation model. *Mechanical Systems and Signal Processing, 45*(2), 396-407.

Tseng, S. T., Hamada, M., & Chiao, C. H. (1995). Using Degradation Data to Improve Fluorescent Lamp Reliability. *Journal of Quality Technology, 27*(4), 363-369.

Upadhyaya, B. R., Naghedolfeizi, M., & Raychaudhuri, B. (1994). Residual life estimation of plant components. *P/PM Technology, 7*(3), 22-29.

Wang, T. (2010). *Trajectory similarity based prediction for remaining useful life estimation.* University of Cincinnati.

Yang, K., & Jeang, A. (1994). Statistical surface roughness checking procedure based on a cutting tool wear model. *Journal of Manufacturing Systems, 13*(1), 1.

Zhang, C., Lim, P., Qin, A. K., & Tan, K. C. (2017). Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. *IEEE Trans Neural Netw Learn Syst, 28*(10), 2306-2318. doi:10.1109/TNNLS.2016.2582798

Zhang, Z., Si, X., Hu, C., & Lei, Y. (2018). Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. *European Journal of Operational Research*.