Monotonic Feature Extraction for Equipment Remaining Life Prediction Using Dynamic Time Warping

1st Wu Xiaotong
College of Automation Engineering , Nanjing University of Aeronautics and Astronautics Nanjing,China
13770221317@163.com

2nd Lu Ningyun
*College of Automation Engineering , Nanjing University of Aeronautics and Astronautics,*Nanjing,China
luningyun@nuaa.edu.cn

3rd Lyu Jianhua
School of Computer Science and Technology, Southeast University
Nanjing, China
lujianhua@seu.edu.cn

***Abstract*—This paper proposes a method to screen out monotonic features for equipment remaining life prediction (RUL) using Dynamic Time Warping (DTW). First, Spearman correlation coefficient is used to pre-screen the original features to build a candidate feature set. Second, trend sequences are constructed from the diagonal regions of DTW cost matrix and they have the same height and width as the feature sequences. Then, the similarity index between the candidate features and the trend sequences is calculated, the features with high similarity are screened out since they have strong correlation with the trend sequences, and all screen-out features form a "core" feature set. Finally, these core features are input to a Long Short Term Memory (LSTM) neural network model to predict the RUL. Experimental verification is carried out using a real-world dataset for low-voltage circuit breakers, and the results show that the features screened out by the proposed method are effective for accurate RUL prediction.**

Keywords—Monotonic feature extraction, Remaining useful life (RUL) prediction, Dynamic Time Warping (DTW)

I. INTRODUCTION

Equipment performance will inevitably deteriorate with the increase of serving time due to comprehensive influences of various internal causes such as aging of component, mechanical wearing and material fatigue, and external causes such as vibration and shock[1]. The operation performance and health status of complex system or equipment will inevitably show a trend of life loss. When the loss reaches a certain level, the equipment is unlikely able to complete normal tasks and functions, resulting in irreversible economic losses and waste of resources [2]. In order to ensure the safety and reliability of the equipment, it is necessary to predict the equipment remaining life before serious accidents occur, especially in the early stage of equipment system performance degradation.

Obviously, there are a large number of factors that affect the degradation process of the equipment. If every possible influencing factor is analyzed, it will not only be time-consuming and labor-intensive, but also introduce redundant influencing factors and reduce the prediction accuracy. Therefore, it is very necessary to extract the right features that have the significant impact to the degradation process of the equipment. To do so, we can reduce the dimension of feature space, and obtain a set of features with as few numbers as possible and as little loss of information as possible; and at the same time, the prediction accuracy can be guaranteed, to improve the efficiency and precision of the equipment remaining life prediction model.

There are already a lot of feature extraction methods. However, some of them highly rely on expert experience, resulting in narrow scope of application. Some use deep learning to extract features, but the extracted features are often lack of interpretability.

This paper is from a monotonicity point of view. In general, the life of the degradation state of the equipment is often monotonic, and the core factors that cause the change of the internal state of the equipment are also monotonic [3],

so the characteristics that can reflect the change of the internal health state of the equipment and can be used for the prediction of the remaining life should also have good monotonicity[4].According to the above analysis, the Dynamic Time Warping (DTW) algorithm is used to screen out the features with better monotonic trend to form a feature set for life prediction [5]. DTW is an algorithm that commonly used to quantify the similarity between different trajectories in various fields[6-12], such as speech recognition, transportation navigation electrocardiogram analysis, and stock analysis, etc.

The proposed monotonic feature extraction method for equipment remaining life prediction (RUL) using DTW includes the four main steps: 1) pre-screen the original features to build a candidate feature set, 2) construct trend sequences that have the same height and width as the feature sequences, 3) calculate the similarity index between the candidate features and the trend sequences to screen out "core" features, 4) build RUL prediction model based on the core features. Experimental verification is carried out using a real-world dataset for low-voltage circuit breakers, and the results show that the features screened out by the proposed method are effective for accurate RUL prediction.

II. MONOTONIC FEATURE EXTRACTION USING DTW

A. DTW Algorithm

Suppose there are two time series $X = \{x_1, x_2, \cdots, x_m\}$ and $Y = \{y_1, y_2, \cdots, y_n\}$, where m and n are the lengths of sequence X and sequence Y respectively. Fig.1 illustrates an example of DTW alignment of two time series. The left is the corresponding relationship diagram of each time point of the two time series, and the black dashed line represents the alignment path between the points. The right shows the cost matrix representation of the two time series, and the gray part represents the optimal alignment path of the two time series.

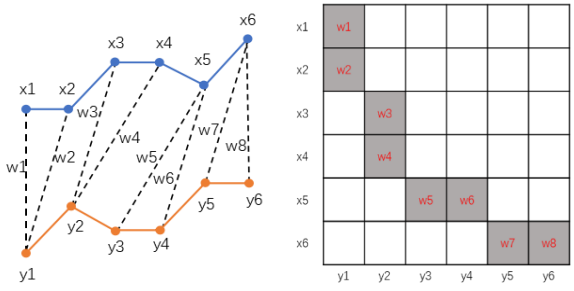


Fig. 1. DTW alignment of two time series

To solve the optimal alignment path of two time series, first calculate the distance between each time point of time series X and each time point of time series Y, and construct the cost matrix T of these two time series,

$$T_{m\times n} = \begin{bmatrix} d(x_1, y_1) & \cdots & d(x_m, y_1) \\ \vdots & \ddots & \vdots \\ d(x_1, y_n) & \cdots & d(x_m, y_n) \end{bmatrix}, \quad (1)$$

where $d(x_m, y_n)$ represents the matching relationship between the time point x_m and the time point y_n.

Then, the dynamic programming idea is used to find the optimal alignment path. The constraints are:

1) The starting point of the path must be $d(x_1, y_1)$, and the end point must be $d(x_m, y_n)$;

2) Elements on the path must follow adjacent points down.

One should find the optimal path among many possible paths using dynamic programming to guarantee the sum of distance elements on the path is the smallest, and then calculate the cumulative distance matrix:

$$D(x_m, y_n) = d(x_m, y_n) + min\begin{Bmatrix} D(x_m, y_{n-1}) \\ D(x_{m-1}, y_{n-1}) \\ D(x_{m-1}, y_n) \end{Bmatrix} \quad (2)$$

Then, find the minimum of the summed path distance elements by:

$$D_{DTW} = min\left(\sum_{i=1}^{k} d_i\right) \quad (3)$$

where, d_i represents the distance element corresponding to the i-th point on a certain path, k represents the total number of elements on the path [13].

B. Monotonic Feature Extraction

Considering the time complexity of the original DTW algorithm, in this paper, the Spearman correlation coefficient is firstly used to calculate the correlations between the original feature set and the equipment remaining life (available in offline modeling). The Spearman coefficient uses the rank of the variable to replace the variable itself to evaluate the trend of the variable. The formula for calculating

the Spearman coefficient between a characteristic variable and the equipment running time is shown in Eq.(4),

$$\rho = \frac{\sum_i (T_i - \bar{T})(X_i - \bar{X})}{\sqrt{\sum_i (T_i - \bar{T})^2 \sum_i (X_i - \bar{X})^2}} \quad (4)$$

In Eq.(4), $\{T_i\}_{i=1:L}$ and $\{X_i\}_{i=1:L}$ represent the order sequences of time $\{t_i\}_{i=1:L}$ and variable $\{x_i\}_{i=1:L}$ respectively; L is the sequence length. $\bar{T}$ and $\bar{X}$ represent the mean of the order columns $\{T_i\}_{i=1:L}$ and $\{X_i\}_{i=1:L}$ respectively.

When the variable shows a monotonically increasing or monotonically decreasing trend with time, the corresponding $|\rho|$ approaches 1. On the contrary, when the variable remains constant over time or fluctuates randomly, the corresponding value of ρ is close to zero. A larger $|\rho|$ indicates that the variable is more related to equipment degradation[14].

Determine the candidate feature set according to the absolute value of the correlation coefficient. Then the DTW algorithm is used to calculate the similarity between the candidate features and the diagonal to measure the monotonicity of the candidate features. The diagonals of the feature sequences are the diagonal sequences of the regions with equal width and height to the feature sequences, and the sequences are linearly related to the RUL.According to the above algorithm, the process of extracting features with monotonic trend using DTW are given as follows.

1) Read in feature sequence.
2) Compute the extreme values of the feature sequence. To build a diagonal of the region with the same heigh-width as the feature sequence, one should calculate the maximum value *max* and the minimum value *min* of the feature sequence as the two endpoints of the diagonal.
3) Create trend sequences. If the length of the feature sequence is *n*, two trend sequences can be created as Table 1 shows. One is an arithmetic descending sequence with an initial value of *max* and a difference value of (*max*-*min*)/*n*, and the other is an equidistant increasing sequence with an initial value of *min* and a difference value of (*max*-*min*)/*n*.
4) Calculate the distance between the feature sequence and the trend sequence A as D_1, and the distance between the feature sequence and trend sequence B as D_2 using DTW. The reciprocal of the minimum value of the two results ρ_{DTW} is used as the criteria for judging the trend of the feature sequence. That is:

$$\rho_{DTW} = \frac{1}{\min(D_1, D_2)} \quad (5)$$

The larger the ρ_{DTW}, the smaller the distance between the feature sequence and the trend sequences, the stronger the correlation between the feature sequence and the trend sequences, and the stronger the monotonicity of the feature sequence.

5) Repeat steps 1) to 4) for each feature sequence, and get a set $\mathbf{P}_{DTW} = \{\rho_{DTW1}, \rho_{DTW2}, \cdots, \rho_{DTWn}\}$with *n* criteria. Sort the elements of the set from largest to smallest, and select the features with the largest criteria according to the requirements to build a set of core features.

TABLE I. TRNED SEQUENCES

	Start	**End**	**Difference Value**
Trend Sequence A	*max*	*min*	(*max*-*min*)/*n*
Trend Sequence B	*min*	*max*	(*max*-*min*)/*n*

C. Predictive Model for RUL

As well know, a neural network with the original fully-connected structure is difficult to deal with the sequence data that are related to each other between the data before and after. Recurrent Neural Network (RNN) can effectively mine sequence features by introducing the concept of hidden state and considering the correlation between the sequence data before and after. However, RNN has the problem of vanishing gradient, which makes it difficult to learn long-term dependency information. In order to avoid gradient disappearance, Long Short Term Memory (LSTM) is used because it is more advantageous in solving long sequence problems. Fig.2 illustrates a typical LSTM neural unit[15].

In the network structure used in this paper, the input layer is first connected to the LSTM layer, then the LSTM layer is connected to 3 fully connected layers, and the last fully connected layer is used as the output layer. The input layer inputs the characteristics of 5 opening and closing operations at a time. In order to avoid overfitting when the

amount of data is small, a dropout layer is added after each layer of the LSTM layer and the fully connected layer, and the neural network unit is temporarily discarded from the network according to a certain probability. The specific network structure is shown in Fig 3.

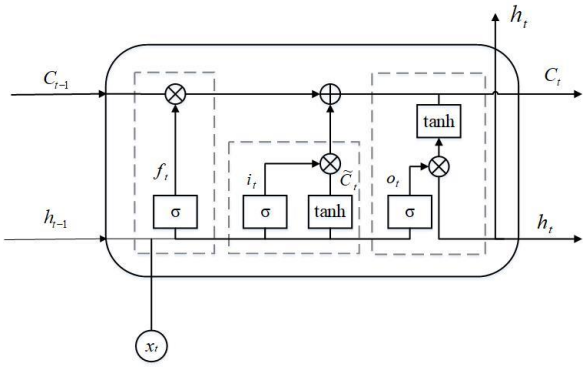


Fig. 2. The basic structure of LSTM recurrent neural unit

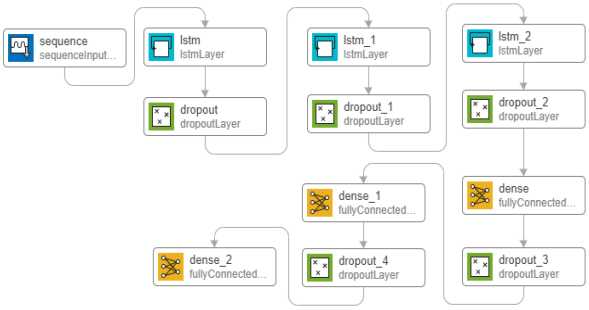


Fig. 3. Prediction Network Structure

D. Method Evaluation Criteria

1) Root mean square error (RMSE)

RMSE is a commonly used performance indicator in regression prediction problems, which can reflect the deviation between the predicted value and the true value. Its calculation formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(L_{real_i} - L_{pre_i})^2} \tag{6}$$

where L_{real_i} is the true value of the *i*-th test sample, L_{pre_i} is the predicted value of the *i*-th test sample. *n* is the number of test samples.

2) Prediction accuracy (Acc)

This paper uses Eq.(7) to evaluate the accuracy of equipment remaining life prediction.

$$Acc = 100\% - \frac{|L_{real} - L_{pre}|}{L_{all}} \tag{7}$$

where Acc is the prediction accuracy, L_{real} is the actual remaining life, L_{pre} is the predicted remaining life, and L_{all} is the equipment life value. Acc ranges from 0 to 1. The closer to 1, the higher the prediction accuracy.

III. EXPERIMENTAL VERTIFICATION

The data set used for the experimental verification in this paper is the accelerated life test data of the same type of low-voltage circuit breakers. The data set includes the full life cycle data of 34 low-voltage circuit breakers, and records the changes of current and voltage of each equipment during each 'closing-opening' operation from brand new to final failure. The low-voltage circuit breakers in the data set were tested in the same accelerated life test platform and experimental conditions. The reason for the failure of the contactor is the failure of fusion welding, that is, due to the action of the arc, the surface of the moving and static contacts melts, causing the moving and static contacts to stick together and cannot be separated, which eventually leads to the failure of the equipment.

First, we extracts domain features, general features and latent features as original features, and labels the features. Then calculate the Spearman correlation coefficients between the original features and the RUL, and determine the candidate features according to the absolute value of the correlation coefficients. After pre-screening, 25 features were screened out. Fig.4 shows some of the results.

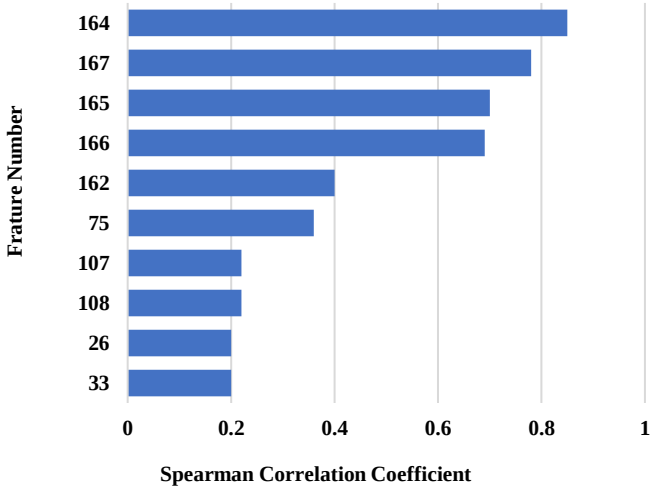


Fig. 4. Spearman Correlation Coefficient Result

Then, according to the steps of using DTW to screen out monotonic features described in Section II, build trend sequences and calculate the correlation between the feature sequence and the trend sequences.

Fig.5 shows part of the results after fitting the feature sequences and trend sequences using DTW, where the blue line is the feature sequences and the black line is the trend sequences. After normalization, the trend criteria of feature 1, feature 2, feature 3, and feature 4 are 0.983, 0.895, 0.657 and 0.378 respectively. The relative magnitude of the trend criterion is consistent with the intuitive visual effect of the feature trend.

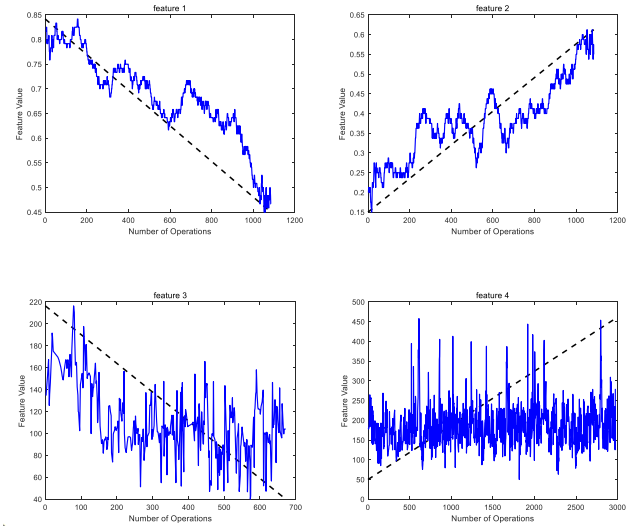


Fig. 5. Results of fitting feature sequences and trend sequences using DTW

For the data set used in this experiment, a total of 5 features with good monotonicity trend were screened out. It is necessary to pay attention to both the long-term trend of the internal state change of the device and the real-time state change. Perform differential operations, complement operations, window statistics, and three-phase combined statistics on the two features with the best trends. Finally got 20 core features.

Use the selected core features with better monotonic trend to predict the life of the equipment.Take 27 devices as training set.The remaining 7 devices are used as the test set. Part of the results is shown in Fig.6.The left column is the prediction curve.The blue line is the predicted result, the yellow line is the true lifespan, and the green line is the more stable result obtained by linear regression of the predicted value.The right column is the prediction accuracy. The red line represents 80% prediction accuracy. The blue line represents the actual prediction accuracy.

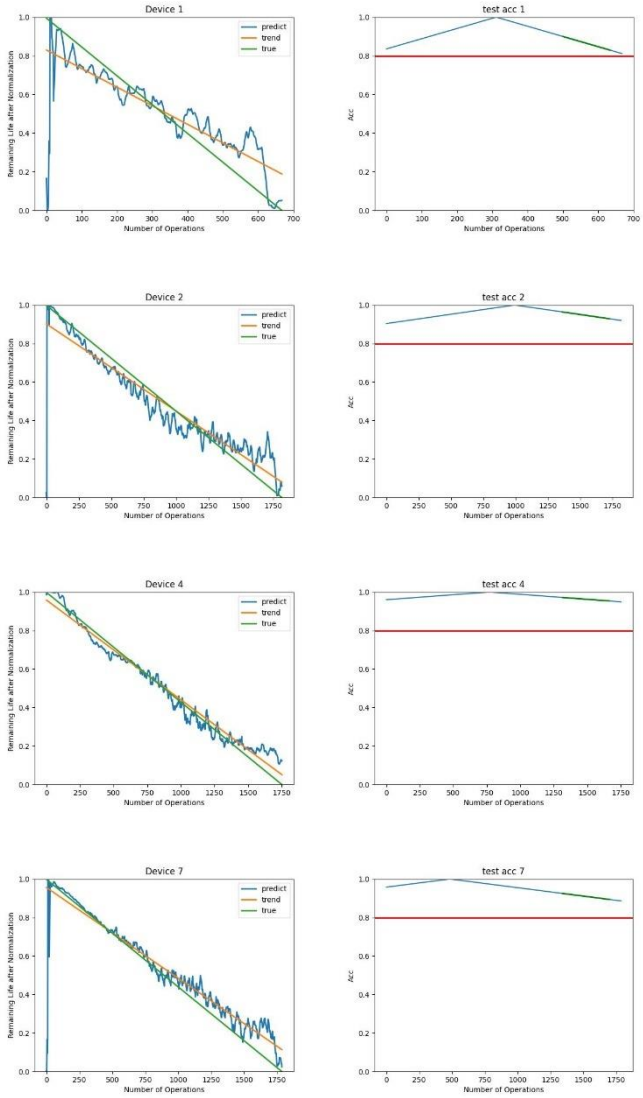


Fig. 6. Prediction results

From the prediction accuracy curve, it can be seen that the prediction accuracy of the remaining life is all above 80% which well meet the requriement of pratical application.

In order to verify the effectiveness of the method in this paper, comparative experiments were carried out. Table 2 lists the RMSE of the prediction results using the original features. Table 3 lists the RMSE of the prediction results using the features extracted by the method in this paper.

It can be seen from the results that the prediction error can be effectively reduced by using the method in this paper.

TABLE II. RMSE RESULTS USING THE ORIGINAL FEATURES

Device ID	1	2	4	7
RMSE	0.404	0.062	0.096	0.179

TABLE III. RMSE RESULTS USING THE EXTRACTED FEATURES

Device ID	1	2	4	7
RMSE	0.095	0.048	0.026	0.058

IV. CONCLUSION

This paper proposes a method to screen out core features with monotonic trends for equipment life prediction using Dynamic Time Warping. In this method, the similarity measure of two time series is realized by using DTW. Two sets of trend sequences are constructed based on the degradation trend of lifespan itself. Make the candidate features to measure the similarity with these two kinds of trend sequences respectively. After sorting, the core features with the strongest monotonic trend are screened out for remaining life prediction. Experiments are carried out using the data set of low-voltage circuit breakers, and the results showed that the prediction accuracy reached more than 80%. How to improve the prediction method to further improve the accuracy is the next research goal.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (62020106003; 61873122).

REFERENCES

[1] CHEN Chuang,LU Ningyun,JIANG Bin,XING Yin.Condition-based maintenance optimization for continuously monitored degrading systems under imperfect maintenance actions[J].Journal of Systems Engineering and Electronics,2020,31(04):841-851.

[2] Pei Hong, Hu Changhua, Si Xiaosheng, Zhang Jianxun, Pang Zhenan, Zhang Peng. A review of methods for predicting remaining life of equipment based on machine learning [J]. Journal of Mechanical Engineering, 2019, 55(08): 1-13.

[3] Zhao Jie, Dong Jimin, Zhang Yanping. Arc fault diagnosis method based on multiple electrical incremental fusion criteria [J]. Electrical Technology, 2022, 23(02): 79-87.

[4] Zhu Kaiqiang, Lu Ningyun, Jiang Bin. Remaining life prediction of high-speed rail traction system based on performance degradation [J]. Information and Control, 2020, 49(03): 335-342. DOI: 10.13976/j.cnki.xk. 2020.9365.

[5] Chen Hao. Research on the failure prediction method of helicopter key components based on deep learning [D].National University of Defense Technology,2019.DOI:10.27052/d.cnki.gzjgu.2019.000737.

[6] Toni Giorgino.Computing and Visualizing Dynamic Time Warping Alignments in R:The dtw Package[J].Journal of Statistical Software,2009,31(7).

[7] Alejandro Gaston Alvarez, Diego, Alexis Evin, Sebastian Verrastro. Implementation of a Speech Recognition System in a DSC[J]. IEEE Latin America Transactions,2016,14(6).

[8] R.Varatharajan, Gunasekaran Manogaran, M.K.Priyan,Revathi Sundarasekar.Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm[J]. Cluster Computing,2018,21(1).

[9] Lauder B, Schwerin B , Mcconnell M , et al. Using Dynamic Time Warping for Noise Robust ECG R-peak Detection[C]// 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS). 2019.

[10] Tian Han, Qin-Ke Peng, Zhi-Bo Zhu, Yi-Qing Shen, Hui-Jun Huang, Nahiyoon Nabeel Abid. A pattern representation of stock time series based on DTW[J]. Physica A: Statistical Mechanics and its Applications,2020,550(prepublish).

[11] Cao Shuangpeng, Xu Xiaofeng, Wang Liang, Xu Dong, Yao Ye. Research on Operation Reliability Evaluation Method of GIS Internal Disconnect Switch Based on DTW Algorithm Waveform Similarity [J]. Journal of Shenyang Institute of Engineering:natural Science,2020,16(04):5660.DOI:10.13888/j.cnki.jsie(ns).2020.04.012

[12] Ye Lilun. Research on the prediction of the remaining service life of CNC machine tools based on deep learning [D]. Xidian University, 2021. DOI: 10.27389/d.cnki.gxadu.2021.000904.

[13] Qiao Meiying,Liu Yuxiang,Tao Hui.A Multidimensional Time Series Similarity Measurement Algorithm Based on Information Entropy and DTW[J].Journal of Sun Yat-Sen University(Natural Science Edition),2019,58(02):18.DOI:10.13471/j.cnki.acta.snus.2019.02.001

[14] Ying Rongrong. Research on Remaining Life Prediction Technology Based on Similarity[D]. Southeast University, 2021. DOI: 10.27014/d.cnki.gdnau.2021.003271.

[15] Chen Ziqiang. Research on equipment health status assessment and remaining life prediction method based on LSTM network[D]. University of Science and Technology of China, 2019.