Method for Predicting Cutter Remaining Life Based on Multi-scale Cyclic Convolutional Network (MSRCNN)

Tao Li¹, YinQuan Yu¹*, JinWen Yang¹, Long Zhang², WenBin Tu² and Hao Yong²

¹Institute of Precision Machining and Intelligent Equipment Manufacturing from School of Mechatronics and vehicle engineering, East China Jiaotong University, Nanchang, Jiangxi, 330013, China
²Mechatronics and vehicle engineering, East China Jiaotong University, Nanchang, Jiangxi, 330013, China
*Corresponding author’s e-mail: yu_yinquan@ecjtu.edu.cn

Abstract: In the process of predicting the remaining cutter life (RUL), the deep learning method such as convolutional neural network (CNN) does not consider the time correlation of different degradation states, which directly affects the accuracy of the remaining cutter life prediction. To extract the features with time series information to predict the RUL more effectively, this paper proposes a new deep neural network, which is named the multi-scale cyclic convolutional neural network (MSRCNN). In the MSRCNN, a cyclic convolutional layer is constructed to simulate the time correlation of different degradation states for mining the timing characteristics of the data. Multi-scale features are extracted through multi-scale convolution, and the convergence of parameters is improved by layer-by-layer training and fine-tuning. Finally, the RUL is predicted based on the features. The comparison with the published prediction methods of CNN and recurrent neural network (RNN) models proves that our proposed method (MSRCNN) is effective and superior in improving the accuracy of RUL prediction.

1. Introduction

In advanced manufacturing systems, the high performance of a machine tool is the key to producing high-quality machined surfaces, and the main cause of cutter failure is cutter contact tip wear. To ensure machining accuracy within the cutter life cycle, the industry generally adopts excessive protection strategies, resulting in additional machining cost and unnecessary cutter change downtime. To develop an effective RUL prediction system, scholars have carried out more and more researches. At present, RUL prediction methods are mainly divided into two categories, namely methods based on physical models and methods based on data-driven [1]-[13].

During tool wear, with the passage of time, the cutter degenerates from normal operation to complete failure, and the development of failure is also a gradual evolution process. Correspondingly, the degradation state of the cutter at different moments is related to the time scale. However, the existing research ignores this dependence in the network construction process, which affects the accuracy of the prediction model and limits its promotion. Therefore, the establishment of correlation models of different degradation states is very important to predict the RUL of the cutter accurately.

To effectively use the timing information which hidden in the signal, this paper proposes a new deep learning method, namely the multi-scale cyclic convolutional neural network (MSRCNN), which is...
adopted to predict the remaining cutter life (RUL).

2. Architecture of MSRCNN Model

2.1 Cyclic convolutional layer

The convolutional layer is the core component of CNN, which can automatically extract discriminative features from the input time sequence sensor data. However, there is no loop layer in the convolutional layer, which cannot feed the output back to the input. It means that information only flows forward in the CNN. Correspondingly, only the current input information in each time step is considered and the previous degradation information is ignored in CNN. In particular, the existing based on CNN cutter RUL prediction methods cannot solve this issue, which reduced their prediction accuracy and generalization ability. Therefore, in this article, a new cyclic convolutional layer is constructed to solve this problem and improve the prediction performance of the network.

As shown in Fig. 1, \( x^t = f(x^{t-1}, x_t) \) where \( f(\cdot) \) is the nonlinear activation function, \( x^{t-1} \) is the input time sequence sensor data, and \( h_{t-1} = x_{t-1} \) is the storage state fed back by the loop connection at time step \( t-1 \).

![Fig. 1 Cyclic convolution layer gate control mechanism](image)

There are two gates in the gated recursive convolutional layer, namely the reset gate \( r^t \) and the update gate \( u^t \) as given by:

\[
    r^t = \delta(K_r^t \ast x^{t-1} + W_r^t \ast h_{t-1} + b_r^t)
\]

\[
    u^t = \delta(K_u^t \ast x^{t-1} + W_u^t \ast h_{t-1} + b_u^t)
\]

where \( \delta(\cdot) \) is the Logistic Sigmoid function, \( \ast \) represents the convolution operator, \( K_r^t, W_r^t, K_u^t, W_u^t \) are the convolution kernels, \( b_r \) and \( b_u \) are the bias terms. At each time step \( t \), the state of the gated recursive convolutional layer \( x^t \) can be obtained by:

\[
    x^t = u^t \circ h_{t-1} + (1 - u^t) \circ \hat{h}^t
\]

\[
    \hat{h}^t = \tanh(K_h^t \ast x^{t-1} + W_h^t \ast (r^t \circ h_{t-1}) + b_h^t)
\]

where \( \hat{h} \) represents the newly generated state, \( \tanh(\cdot) \) is the activation function, \( K_h^t \) and \( W_h^t \) are the convolution kernels, \( b_h \) is the bias term, and \( \circ \) represents the Hadamard product (matrix elements correspond to multiplication). It can be seen from the equation (3) that the state at time step \( t \), \( x^t \) is a linear combination of the previous state and the current candidate state, which is controlled by the reset and update gates.

2.2 Multi-scale and one-dimensional sensing data

The input data of MSRCNN is various parameters collected by multiple sensors. To comprehensively utilize all the sensor data, this paper uses the sliding window strategy to construct multi-channel one-dimensional sensor data. The process can be expressed as:
\[ I(i, w, t) = x_i(i : w + i) \]
\[ i = 1, 2, ..., m - w \]

among them, \( I \) is the constructed multi-channel data, \( w \) is the length of the sliding window, \( m \) is the lifespan, \( t \) is the number of corresponding channels represents different sensing parameters.

### 2.3 The overall layout of MSRCNN

The structure of the proposed MSRCNN is shown in Fig. 2. In the proposed MSRCNN, the multi-scale cyclic convolutional layer (MSRCL), the pooling layer (PL), and the fully connected layer (FCL) are shown in Fig. 2, respectively. To integrate the degradation information from different sensors, this paper uses the size of the multichannel time series sensor data as the input of the prediction network, where \( H \) is the length of each sensor sequence and \( C \) is the number of sensors. Then, using \( N \) recursive convolutional layers and \( N \) pooling layers to automatically extract data features, and the temporal correlation of different degradation states is modeled. For the recursive convolution layer, where \( i = 1, 2, 3, ..., N \), all convolution kernels have the same parameter settings, that is, the number of kernels is set to \( 2^iM \), and the kernel size is \( 1 \times K \). For the first \( 1-N \) pooling layers, maximum pooling is used as the Subsampling function, and non-overlapping windows are adopted to perform the pooling operation. The final pooling layer, the \( N \)th pooling layer, uses global maximum pooling for Subsampling, and accordingly, the high-level representation from the number \( N \) recursive convolutional layer is converted into a vector of size \( 2^N-1M \). After that, the vector is input into the subsequent \( L \) fully connected layers to perform RUL estimation. In this article, \( L \) is set to 3. In other words, there are 3 fully connected layers in MSRCNN. The first two fully connected layers have \( F \) neurons, and they both use Rectified Linear Unit (ReLU) to achieve nonlinear activation. The third fully connected layer has only one neuron, which serves as the output layer of MSRCNN to predict RUL.

![Fig.2 Architecture of multi-scale cyclic convolution network](image)

### 3. Experimental setup and data processing

#### 3.1 Experimental setup

The installation of a CNC milling machine and sensor is shown in Fig. 3. The artifacts are cut and removed from the raw material, face milled to remove the original skin layer containing hard particles, and then the surface is machined. A Kistler9265B three-way dynamometer is installed between the artifacts and the processing test bench to measure the cutting force in the form of electric charge and convert it into a voltage through the Kistler5019A charge amplifier. Three Kistler piezoelectric accelerometers are installed on the artifacts to measure the vibration of the machine tool in the \( X \), \( Y \), and \( Z \) directions. An acoustic emission (AE) sensor is installed on the artifacts to monitor the high-frequency stress waves generated during the cutting process. Acquisition card NI DAQ data acquisition card collected 7 signals (force_\( x \), force_\( y \), force_\( z \), acce_\( x \), acce_\( y \), acce_\( z \), AE). Select National Instruments LabVIEW 8.2 to create a user-friendly graphical user interface (GUI).
3.2 Data description

To further prove the effectiveness and superiority of the proposed method in the remaining service life prediction, this section applies MSRCNN to the RUL prediction of milling cutters. The experimental data comes from the New York Society for Forecasting and Health Management (New York Society for Forecasting and Health Management). PHM shares the data for the 2010 high-speed CNC machine cutter health prediction competition.

3.3 Evaluation Index

To evaluate the performance of the proposed method in RUL prediction, this paper uses four evaluation indicators as follows: 1) Explained variance score (EVS), explained the variance score of the regression model. Its value range is [0, 1]. The closer it is to 1, the more the independent variable can explain the variance of the dependent variable. The smaller the value, the worse the effect. 2) Mean absolute error (MAE), used to evaluate how close the predicted result is to the real data set. The smaller the value, the better the fitting effect. 3) Mean squared error (MSE), this indicator calculates the mean value of the sum of squared errors between the fitting data and the corresponding sample points of the original data. The smaller the value, the better the fitting effect. 4) R2 score ($R^2$) coefficient of determination, its meaning is also to explain the variance score of the regression model. Its value range is [0, 1]. The closer to 1, the more the independent variable can explain the variance of the dependent variable. The smaller the value, the more that means the worse the effect. Where $y$ is the true value, $\hat{y}$ is the predicted value, and $\bar{y}$ the mean value.

$$\text{EVS}(y, \hat{y}) = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} = 1 - \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

4. Results and discussion

In the RUL prediction, this paper performs 5 cross-validations on the training data set to determine the structural parameters of MSRCNN, including the number of convolution kernels $M$, the kernel size $1 \times K$, the number of recursive convolution layers $N$, the pool size $p$, and the number of neurons $F$. At the same time, this paper applies dropout and L2 regularization to each recursive convolutional layer and each fully connected layer and performs $V$ random forward pass to obtain the predicted mean and variance. Besides, this paper uses the mean square error as the loss function of RCNN and uses the Adam optimizer to optimize the loss of the objective function in the equation by iteratively updating the
network weights and deviations with a minimum batch size of 128.

The MSRCNN model is mainly composed of the CNN and the RNN. To verify the effectiveness of the model, the model is compared with the CNN model and the RNN model. Taking the two sets of cutter data of milling cutter 1 and milling cutter 2 as the test set, the above three models are used to predict their remaining service life respectively. The prediction results are shown in Fig. 7, where the x-label is time, and the y-label is the percentage of the remaining life corresponding to the sample at the current moment in the life cycle. The blue and green, brown solid lines represent the predicted life values, and the red solid lines are the actual life values, respectively.

![Fig. 4 Prediction results of milling cutter RUL with different models](image)

(a) Comparison between CNN model and MSRCNN model (Cutter1); (b) Comparison between RNN model and MSRCNN model (Cutter1); (c) Comparison between CNN model and MSRCNN model (Cutter2); (d) Comparison between RNN model and MSRCNN model (Cutter2).

Fig. 4 shows that the goodness of fit of the MSRCNN model is best compared to the other two models, and are closest to the real RUL.

| The test cutter | The evaluation index | CNN     | RNN     | MRCNN   |
|-----------------|----------------------|---------|---------|---------|
| Cutter 1        | EVS                  | 0.858   | 0.857   | 0.953   |
|                 | MAE                  | 0.101   | 0.141   | 0.051   |
|                 | $R^2$                | 0.911   | 0.860   | 0.950   |
|                 | MSE                  | 0.088   | 0.094   | 0.004   |
| Cutter 2        | EVS                  | 0.846   | 0.838   | 0.947   |
|                 | MAE                  | 0.109   | 0.151   | 0.054   |
|                 | $R^2$                | 0.908   | 0.828   | 0.928   |
|                 | MSE                  | 0.089   | 0.097   | 0.006   |

It can be seen in Table 1 that the MSRCNN model proposed in this article is better than CNN and RNN in the evaluation indicators EVS, MAE, $R^2$, and MSE, which shows that MSRCNN can provide more accurate RUL prediction results, and its performance is steadier. The comparison results show that
the MSRCNN model is conducive to the improvement of accuracy and further improves the accuracy of RUL. Due to the powerful feature extraction ability of multi-scale convolution and the ability of circular convolution to excavate the time sequence characteristics of data, the introduction of time sequence features can reduce the prediction error.

5. Conclusion
In this article, a new prediction framework MSRCNN for RUL prediction of cutting cutters is proposed. The proposed neural network takes the time series data collected by different sensors as input and constructs a cyclic convolution layer to simulate the degraded state of the cutter to mine the time series characteristics of the data. Then, multi-scale features are extracted through multi-scale convolution, and the parameters are optimized by layer-by-layer training and fine-tuning. The maximum pooling layer is used to reduce the representation dimension and make the extracted features more compact. By periodically superimposing multiple cyclic convolutional layers and maximum pooling layers, feature information is automatically extracted from the input data. Finally, by inputting these learned features into the subsequent fully connected layer to estimate RUL, CNN and RNN prediction methods are compared. Experimental results show that compared with the existing CNN-based prediction models and RNN-based prediction models, the proposed MSRCNN has obvious advantages in accuracy and convergence. We can conclude that our proposed method breaks through the limitations of the convolutional neural network prediction model in the evaluation of the remaining service life of the cutter.

Acknowledgment
This research is supported in part by the East China Jiao Tong University Fundamental Research Funds under Grant 2003419018; by the National Natural Science Foundation of China (51965018).

Reference
[1] X. Li, H. Dong, Z. Wang, F. Han, Set-membership filtering for state-saturated systems with mixed time-delays under weighted try-once-discard protocol, IEEE Transactions on Circuits and Systems II: Express Briefs, 66 (2019) 312-316.
[2] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, IEEE Transactions on Reliability, (2018) 1-12.
[3] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Health assessment and life prediction of cutting cutters based on support vector regression, J. Intell. Manuf.26 (2015) 213–223.
[4] D. Wu, C. Jennings, J. Terpenny, R.X. Gao, S. Kumara, A comparative study on machine learning algorithms for smart manufacturing: cutter wear prediction using random forests, Journal of Manufacturing Science and Engineering, 139 (2017) 1-9.
[5] C. Drouillet, J. Karandikar, C. Nath, A.-C. Journeaux, M. El Mansori, T. Kurfess, Cutter life predictions in milling using spindle power with the neural network technique, J. Manuf. Processes 22 (2016) 161–168.
[6] J. Yan, J. Lee, A hybrid method for on-line performance assessment and life prediction in drilling operations, IEEE Int. Conf. Autom. Logist. 2007 (2007) 2500–2505.
[7] B. Wang, Y. Lei, N. Li, T. Yan, Deep separable convolutional network for remaining useful life prediction of machinery, Mechanical Systems and Signal Processing, 134 (2019) 1-18.
[8] L. Guo, Y. Lei, N. Li, T. Yan, N. Li, Machinery health indicator construction based on convolutional neural networks considering trend burr, Neurocomputing, 292 (2018) 142-150.
[9] Babu GS, Zhao PL and Li XL, Deep convolutional neural network based regression approach for estimation of remaining useful life[J]. Lecture Notes in Computer Science, 2016, 214-228.
[10] Li X, Ding Q and Sun JQ, Remaining useful life estimation in prognostics using deep convolution neural networks[J]. Reliability Engineering and System Safety, 2018, 172:1-11.
[11] Heimes FO, Recurrent neural networks for remaining useful life estimation[C]// International Conference on Prognostics and Health Management (PHM), 2008:1–6.
[12] L. Wen, Y. Dong, L. Gao, A new ensemble residual convolutional neural network for remaining useful life estimation, Mathematical Biosciences and Engineering, 16 (2019) 862-880.

[13] X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, Reliability Engineering & System Safety, 182 (2019) 208-218.