Tool wear condition monitoring in milling process based on data fusion enhanced long short-term memory network under different cutting conditions

Guoxiao Zheng, Weifang Sun, Hao Zhang, Yuqing Zhou,*, Chen Gao

Abstract

Tool wear condition monitoring (TCM) is essential for milling process to ensure the machining quality, and the long short-term memory network (LSTM) is a good choice for predicting tool wear value. However, the robustness of LSTM-based method is poor when cutting condition changes. A novel method based on data fusion enhanced LSTM is proposed to estimate tool wear value under different cutting conditions. Firstly, vibration time series signals collected from milling process are transformed to feature space through empirical mode decomposition, variational mode decomposition and fourier synchro squeezed transform. And then few feature series are selected by neighborhood component analysis to reduce dimension of the signal features. Finally, these selected feature series are input to train the bidirectional LSTM network and estimate tool wear value. Applications of the proposed method to milling TCM experiments demonstrate it outperforms significantly SVR-based and RNN-based methods under different cutting conditions.

Keywords

tool wear condition monitoring, empirical mode decomposition, variational mode decomposition, fourier synchro squeezed transform, neighborhood component analysis, long short-term memory network.

1. Introduction

In the modern numerical control milling process, tool condition is one of the key factors affecting the machining quality of workpiece [19, 22]. Tool breakage is the main cause of abnormal shutdown and lead to time lost and capital destroyed [27]. It has reported that severe tool failure causes at least 20% of abnormal downtime [4, 32]. However, traditional tool condition monitoring (TCM) methods are based on the machining time or the number of workpiece machined resulting in the effective utilization rate of tool is only 50%-80%, which affect the processing efficiency and increase the machining cost significantly [15, 35]. It is predicted that an effective TCM method can increase the cutting efficiency by 10-50% and reduce the machining cost by 10-40% [23, 33]. Therefore, the development of effective online TCM method has received broadly positive reviews and is a research hotspot nowadays [10, 11].

Recently, many deep learning models have been employed in TCM applications [9, 14, 21]. For example, Cao et al [1] recognized tool wear condition by derived wavelet frames and Convolutional neural network (CNN) using vibration signals. Recent advanced technology that have greatly increased the number of TCM study, Huang et al [8] proposed a tool wear predicting method by deep CNN, in which multi-domain features are respectively extracted from cutting force and vibration. Lei et al [16] employed Extreme learning machine (ELM) to classify tool wear condition in milling processes, and used genetic algorithm and particle swarm optimization to optimize model parameters of ELM. Tim and Chris [26] proposed a disentangled-variation-autoencoder CNN method to estimate tool wear condition in a self-supervised way. Zhi et al [30] proposed a hybrid CNN and edge-labeling graph neural network (EGNN) method for limited tool wear image training samples, in which the CNN is employed to extract features of tool wear image and the EGNN is applied to distinguish the tool’s category. However, these TCM methods have been generally applied for diagnosis (classification) rather than prognosis (regression), tool wear is a progressive and continuous cumulative process, regressive prediction of tool wear is more suitable than classification that the CNN-based methods are difficult to use [34]. Recurrent neural networks (RNN) could be solve the problem

(*) Corresponding author.
E-mail addresses: G. Zheng - 900936695@qq.com, W. Sun - vincent_suen@126.com, H. Zhang - jiyongxing@customs.gov.cn, Y. Zhou - zouyq@wzu.edu.cn, C. Gao - gaochen_1993@163.com
of regression and increase the accuracy of the prognosis, but the error of backpropagation in RNN would increase sharply or decrease exponentially, which lead to the problem of long lag [5,18]. As a significant branch of RNN, Long short-term memory (LSTM) network is proposed to overcome the above problem. Due to the special unit structure with learning long-term dependencies, LSTM can deal with the long-distance dependence problem in time sequence data [6]. Therefore, LSTM is potential to obtain good performance for TCM [31], Tao et al [24] designed a TCM method based on LSTM and hidden Markov model (HMM) to estimate the tool wear value and predict it’s remaining useful life. Zhao et al [29] proposed a convolutional Bi-directional LSTM network, in which CNN extracted local feature of original signal and Bi-directional LSTM encoded temporal information and predict tool wear value. However, it is found that the regression accuracies of LSTM- based TCM method are poor when the cutting conditions of testing samples are different with that of training samples in our experiment. That is, the cutting condition could affect significantly the performance of LSTM- based TCM method. Therefore, this paper try to alleviated the influence of cutting condition to LSTM model through a data fusion way.

In this paper, a data fusion enhanced LSTM- based TCM method is established to estimate tool wear value under variable cutting conditions. The paper is organized as follows: Section 2 introduces the proposed data fusion enhanced LSTM method, Section 3 describes the experimental setup, data analysis and experimental results. Finally, conclusion is in Section 4.

2. Proposed method

2.1. Framework of the proposed method

The proposed TCM method framework based on data fusion enhanced LSTM is illustrated in Figure 1. Firstly, vibration time series signal collected from milling process are transformed to feature space through Empirical mode decomposition (EMD), Variational mode decomposition (VMD) and Fourier synchro squeezed transform (FSST), and then few feature series are selected by neighborhood component analysis (NCA) to reduce dimension of the signal features. Finally, these new feature series selected by NCA are input into bidirectional LSTM network to train the regression model.

![Fig. 1. Framework of the data fusion enhanced LSTM- based TCM method](image)

2.3. Feature extraction

2.3.1. Empirical mode decomposition

EMD is a nonlinear time-frequency decomposition algorithm that decompose the signal into several intrinsic mode functions (IMFs) and a residual [7], shown in Equation (4). In EMD, all decomposed IMFs contain the local feature information in different time scales of the original signal. Finally, each IMF contains approximately a single frequency component, and the instantaneous frequency of the original signal can be obtained after the weighted average of the instantaneous frequency of each IMF:

\[ X(t) = \sum_{i=1}^{N} \text{IMF}_i(t) + r_N(t) \]  

(4)

EMD decomposes the signal according to the time scale features of the original data, without pre-setting any basis function, which is the most significant advantage compared with other time-frequency decomposition methods, such as wavelet transform. Due to the complexity and uncertainty of milling process, it is very difficult to find a basis function suitable for milling signal, EMD could be employed for feature extraction in milling TCM.

2.3.2. Variational mode decomposition

VMD is an adaptive time-frequency signal decomposition algorithm, its framework is the solution of variational problems [3]. VMD considers the signal is composed of sub signals with different frequencies dominant, and transforms the decomposition of signal into the solution of constrained variational model [13,28]. In this process, the central frequency and bandwidth of each IMF are updated alternately and iteratively. Finally, the signal band is decomposed adaptively and obtain the preset K narrowband IMFs in equation (5).

\[ x(t) = \sum_{k=1}^{K} u_k(t) \]  

(5)

In VMD, each IMF \( u_k \) is a bandwidth limited frequency modulation and amplitude modulation signal shown in equation (6):

\[ u_k(t) = A_k(t) \cos(\phi_k(t)) \]  

(6)

VMD has perfect mathematical theory support, its essence is an adaptive optimal Wiener filter group, which can get high signal-to-noise ratio IMFs.
The transform values are then “squeezed” so that they concentrate around curves of instantaneous frequency in the time-frequency plane.

### 2.4. Long short-term memory network

An LSTM network is a type of RNN that can learn long-term dependencies between time steps of sequence data [6,29]. The framework of LSTM is shown in Figure 2.

Let \( X_t = \{ X_{t1}, X_{t2}, X_{tC} \} \) is a time series with \( C \) features, \( h_t \) and \( c_t \) are the hidden state and cell state at time \( t \), respectively. At time \( t \), the state of the network \(( c_t, h_t) \) is calculated by \( X_t \) and \(( c_{t-1}, h_{t-1}) \) by Equation (8) and (9):

\[
\begin{align*}
  c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
  h_t &= \sigma_t \odot \sigma_c (c_t) 
\end{align*}
\]

The definition and expression of \( i, f, g, o \) are as shown in Table 1.

### 3. Experimental observation and research

#### 3.1. Experimental setup

The experimental device for milling TCM is shown in Figure 3. In the milling TCM experiment, a CNC milling machine (DMTG VDL850A, China) is used to finish milling process, and a piece of #45 steel (30 cm × 10 cm × 8 cm) is used as the workpiece material. What’s more, the milling vibration signals of spindle X and Y directions are acquired by two accelerometers with a signal acquisition device (ECON Dynamic Signal Analyzer, shown in Figure 3(b)). In addition, the signal sampling frequency in the experiment is 12KHz. Fourteen uncoated three-insert tungsten steel end milling cutters with diameter of 10 mm are employed to mill the workpiece under different cutting conditions, listed in Table 2. For each tool, the tool wear value is measured after milling each surface using a tool microscope (GP-300C Figure 3(c)). The length of rake face wear (KB) is employed as the tool wear criterion in the experiment, and the max value \( KB = \text{max} (KB_1, KB_2, KB_3) \) of three inserts is adopted as the final tool wear value.

Figure 4 illustrates the tool wear conditions after milling the workpiece surface 1-st, 5-th and 10-th times.

In the 14 milling TCM experiments, the training, verification and testing sets are generated randomly shown in Table 2, 7 sets of samples for training, 3 sets of samples for verification, and 4 sets of samples for testing.

### 3.2. Results and analysis

#### 3.2.1. Samples and metrics

Acceleration signals of Spindle X and Y direction are used in the network, 272 training set, 120 validations set, and 80 test set are made up from spindle sensor signals. In all analyzed samples, there is no same cutting condition combination in the three dataset. Besides, in the signal pre-processing, the original signal of each sample is divided into 10 parts by slide window method, in which the window size is 2000 points, and the sliding distance is 1000 points.

To evaluate the performance of the proposed method, three indexes are employed, including the mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R²).

### Table 1. Definition and expression of the LSTM layer gate

<table>
<thead>
<tr>
<th>Component</th>
<th>Purpose</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input gate ((i))</td>
<td>Control level of cell state update</td>
<td>( i_t = \sigma_g (W_i x_t + R_i h_{t-1} + b_i) )</td>
</tr>
<tr>
<td>Forget gate ((f))</td>
<td>Control level of cell state reset (forget)</td>
<td>( f_t = \sigma_g (W_f x_t + R_f h_{t-1} + b_f) )</td>
</tr>
<tr>
<td>Cell candidate ((g))</td>
<td>Add information to cell state</td>
<td>( g_t = \sigma_c (W_g x_t + R_g h_{t-1} + b_g) )</td>
</tr>
<tr>
<td>Output gate ((o))</td>
<td>Control level of cell state added to hidden state</td>
<td>( o_t = \sigma_g (W_o x_t + R_o h_{t-1} + b_o) )</td>
</tr>
</tbody>
</table>

where \( W_i, R_i \) are the input weights and recurrent weight in the t-th layer, and \( b_i \) is the bias of each component.
3.2.2. Algorithm settings

For each cutting process in the experiment, there are two mutually perpendicular milling vibration signals which are collected from the equipment and a part of collected signal has 12000 points as shown in Figure 5, in which the corresponding cutting parameters is the Case 1 of Table 2: spindle speed is 2300 rpm, axial cutting depth is 4 mm, and feed rate is 400 mm/min.

Since the real monitoring signal is often nonlinear and non-stationary, it is suitable to use the EMD, VMD and FSST methods to obtain the features of vibration signals for tool wear. In order to obtain signal...
In this work, the first 6 IMFs and residuals are taken in EMD, the first 5 IMFs and residuals taken in VMD, and 60 IMFs is decomposed in FSST. In addition, it is necessary to take the real and imaginary parts of the IMF as the feature matrix of the vibration signal, and use NCA to take the effective characteristic matrix. The results are listed in the Table 3.

By calculating, it was found that the feature matrix has first 6 numbers of IMFs and residuals of EMD, 5 numbers of IMFs and residuals of VMD, 17 real parts and 14 imaginary parts of FSST. Totally 45 feature matrixes as input signals. The two single-channel experimental data of the sensor are superimposed and fused into a new sample. Meanwhile, all data from experiments need batch normalization.

In this model, it is a way to use eleven layers as neural network architectures in our experiments: especially bidirectional LSTM layer, which has two hidden LSTM layers (forwards and backwards) as shown in Table 4.

Due to the limitations of experimental equipment conditions and cost, 14 sets of experiments were executed, 7 sets of samples under different working conditions were selected for training, 3 sets of samples were selected for verification, and 4 sets of samples were selected for testing.

For all architectures, complete error gradient was calculated and the weights are trained by using gradient descent with momentum. In all experiments, the same training parameters were kept: randomly assigned initial weights, keeping the training algorithm and parameters constant, allowing us to focus on the impact of changing the architecture.

3.2.3. Experimental results

The LSTM model established by the training set and verification set is applied to predict the testing set, including 4 tools with different cutting conditions. In Figure 6, the blue, green, and cerulean dotted lines denote the prediction results using the proposed method with the spindle vibration signal of X-direction, Y-direction, and dual-direction (composition of X and Y directions). It is noted that the cutting parameters of the 5-th and 8-th tools are different. For Figure 6(a), the spindle speed is 2400 rpm, the axial cutting depth is 0.6 mm, and the feed rate is 500 mm/min. For Figure 6(b), the spindle speed is 2500 rpm, the axial cutting depth is 0.5 mm, and the feed rate is 400 mm/min. It can be seen that the trend of the overall predicted value is similar the actual wear value, and the error at some stages is less than 0.1 or even close to the wear value.

To test the regression performance, the proposed method is compared with RNN and support vector machine (SVR). As a result, the MSE, RMSE and $R^2$ of three methods are presented in Table 5.

It can be seen from Table 5 that the proposed LSTM-based method is highly effective in improving the regression accuracy, the prediction accuracy of the proposed method is much higher than that of RNN and SVR according to the values of three evaluation indexes, except for the X-direction signal of the 3-rd and 5-th tools. In addition, the prediction accuracies with the dual-direction signal outperform that of signal-direction except for the 3-rd tool, while the results of three indexes are slightly worse than that of two other methods in the 3-rd tool.


