Application Research on Extreme Learning Machine in Rapid Detection of Tool Wear in Machine Tools

Q Wu¹, E Liu¹, Y H He² and X Tang³

¹School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China
²Information Center, Chongqing University of Technology, Chongqing 400054, China
³DINGZHAO (Chongqing) Packaging Technology Co., Ltd., Chongqing 400050, China
Email: wuqian80@163.com

Abstract. In order to put an end to the product quality accidents caused by cutting tool breakage or severe wear in machining process, the paper explored an ELM model detection method based on voice recognition. In this paper, firstly it analysed the features of cutting sound signal in time-frequency domain, then discussed the extraction method of tool working state sensitive spectral energy statistical feature based on wavelet packet decomposition, and finally constructed an ELM fast detection model based on sound feature recognition. The experimental results demonstrated that the ELM detection model could achieve higher detection accuracy and faster response time. The simulation results show that the ELM model is effective and applicable in detecting tool wear with the help of sound recognition.

Keywords. Cutting tool damage; voice recognition; time-frequency characteristics; wavelet packet decomposition; ELM model.

1. Introduction
In the process of high-speed mechanical finishing, due to the tool tip and the cutting edge are under a lot of mechanical shock, vibration and high temperature and high pressure, such as composite load, easy to cause accidents, and the heavy may lead to machine tool damage and interrupt the entire production line work and bring incalculable loss, therefore, how to timely and accurately identify the running state of the tool has become a research hotspot in the field of intelligent machinery manufacturing. So far, despite many kinds of observation tool condition detection method of geometrical variation, whether direct or indirect detection method, there are more or less certain deficiencies or defects, some even need production line downtime to test the tool condition, the cost is relatively high, Therefore, its detection methods are not always desirable [1-3]. Noteworthy phenomenon is that a cutting operation very experienced skilled operators, often only on the sound signal derived from the cutting region through the auditory organ intuition of the ear, can fairly and accurately distinguish whether cutting tool operation is in normal state, or distinguish the cutting tool wear and breakage degree. This is because skilled operators after long-term training and learning, the auditory system composed of ear perception system and brain neural network is very sensitive to the sound spectrum energy, and its auditory system also shows a strong robustness to the slight changes in the sound spectrum energy. Inspired by the above auditory system, this paper attempts to adopt the model of Extreme Learning Machine (ELM) to imitate skilled operators to identify the acoustic energy
spectrogram characteristics of cutting tools [4], and implement the detection of possible wear state of machining tools with the help of artificial intelligence method.

2. Feature Extraction of Tool Cutting Sound Signal

2.1. Analysis of Acoustic Energy Spectrum Feature of Tool State

The cutting tool sound is a non-speech signal, which expresses a non-stationary process. The focus of people’s attention is to analyze the signal in time domain, frequency domain and power spectrum, in order to find out the characteristic quantity which is strongly related to the working state of the tool, and establish the mapping relationship between the characteristic parameters in time domain and the working state of the tool, so as to determine the characteristic vector of the parameter representing the working state of the tool. Comparing between Fourier transform and Wavelet analysis [5], the latter can achieve the subdivision of low frequency characteristics and high frequency time characteristics of the signal in both time domain and frequency domain, which is more conducive to the extraction of characteristic vectors reflecting the working state of the tool. In particular, the wavelet packet analysis can carry out multi-resolution analysis in both low frequency region and high frequency region of the signal, therefore, it is an analysis method that can better analyze local time-frequency characteristics from both high and low frequency regions [6, 7]. On the condition that the frequency range of the original cutting sound signal is \([0, f_{\text{max}}]\), with a layer of wavelet decomposition, the sound signal can be fully expressed on the approximate part of the frequency band \([0, f_{\text{max}}/2]\) and the detail part of the frequency band \([f_{\text{max}}/2, f_{\text{max}}]\), respectively. In the same way, after the signal has been decomposed by N layer wavelet packet, the signal expression on the frequency band \([0, f_{\text{max}}/2^n]\) can be obtained in the approximate part of the sound signal, and the detail part on the frequency band \([f_{\text{max}}/2^n, f_{\text{max}}/2^{n-1}]\). For example, the schematic diagram of a five-layer wavelet packet decomposition is shown in figure 1.

![Wavelet packet decomposition diagram](image)

**Figure 1.** Five-layer wavelet packet decomposition.

In the figure, S, L and H respectively represent the cutting sound signal, expression of low frequency and high frequency characteristics after signal decomposition. The serial number at the end
of the figure represents the number of decomposition layers, where figures 1a, 1b and 1c are wavelet packet decomposition of high and low frequency decomposition, L1 and H1 approximation and detail characteristics, respectively. The sum of the length of the two parts obtained by each decomposition is the length of the signal of the upper layer. After multi-layer decomposition of the signal, the sum of the signal length of each part of the last layer is the total length of the original signal. The advantage of using the wavelet packet decomposition method is that this method can make the cutting sound signal express in each small frequency band without losing any information of the original signal. Therefore, its accuracy is very high, so far, no other method can match it.

2.2. Feature Extraction of Sound Signal Based on Wavelet Packet Decomposition

Based on the decomposition of N-layer wavelet packet of cutting sound signal, the wavelet packet coefficients of the signal in each frequency segment of the N-layer can be directly calculated. This coefficient determines the frequency energy value of each frequency segment of the original signal. With the help of the wavelet packet decomposition tree as shown in figure 1, the node information of each frequency segment in the decomposition part of the signal in the Nth layer can be read, so as to complete the extraction of the energy spectrum features of sound signals. What is particularly noteworthy here is that the order of nodes represents the time-domain information, namely the order of frequency changes in each small frequency segment. After obtaining the wavelet packet coefficients of the sound signal from low to high frequency, the energy occupied by each frequency band can be calculated by equation (1).

\[ E_{i,j} = \int |S_{i,j}(t)| dt = \sum_{k=1}^{n} |d_{j,k}|^2 \]  (1)

In equation (1), \( E_{i,j} \) represents the energy of the jth node in the ith layer, where i and j are integers, \( d_{j,k} \) (\( k = 1,2,...,n \)) represents the wavelet coefficients of \( S_{i,j} \).

After calculating the energy of each frequency band, the percentage of the energy of each frequency band in the total energy of original signal can be obtained according to equation (2).

\[ L_{i,j} = \frac{E_{i,j}}{E}, \quad E = \sum_{j=1}^{m} E_{i,j} \]  (2)

3. Tool State Recognition Model Based on ELM

3.1. Machine Learning Model

In the machine learning model shown in figure 2, x and y are the sampling samples of the input and output of the actual system respectively, and the learning sample set is \( \{ x, y \} \). The so-called machine learning refers to finding out the predicted output of the actual system under the condition of minimum expected risk.

Reference [8] pointed out that ELM is a new type of neural network. A large number of experiments have proved that ELM is a simple and effective single-hidden layer feedforward neural network learning algorithm. It is only necessary to set the weight number of the hidden layer of the network without adjusting the input weight W of the network and the offset set of the hidden layer.
neurons. It has the advantages of easy parameter selection, fast learning speed, good generalization performance, strong learning ability, and can approximate arbitrary complex continuous function. The basic information processing unit in this network is the neuron.

3.2. ELM Algorithm Model

ELM has gained wide attention in the field of machine learning and artificial intelligence, and has been gradually extended to human behavior recognition, fault diagnosis and other fields. An ELM structure containing M hidden layer nodes is shown in figure 3.

In figure 3, activation function $G(a_i, b_j, x)$ is selected as Sigmoid function, $a_i$ and $b_j$ are connection weights and offset set between input and hidden layer respectively, and $\beta_{L \times M}$ is the weight matrix between hidden layer and output layer. The weights between the input layer and the hidden layer are generated randomly, and there are full connections between the input layer and the hidden layer as well as between the hidden layer and the output layer. N is the number of sampling samples, $x_i$ is the sampling data, and $t_j$ is the classification label number. The relationship between them satisfies equation (3).

$$ \{ (x_i, t_j) \}_{j=1}^{N} \subseteq R^n \times R^n $$

ELM calculation is divided into two steps.
Firstly, for all the training samples collected in the experiment, the output matrix $H$ of the hidden layer node is calculated according to equation (4).

$$ H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_1, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_N, b_1, x_N) & \cdots & G(a_N, b_L, x_N) \end{bmatrix} $$

Then, the least square method is used to calculate the optimal output weight matrix of hidden layer nodes according to equation (5).

$$ \| H \hat{\beta} - Y \| = \min_{\beta} \| H \beta - Y \| $$

When the number of samples N is greater than or equal to the number of hidden layer nodes L, the optimal output weight matrix is calculated according to equation (6).

$$ \hat{\beta} = \left( \frac{1}{C} + H^T H \right)^{-1} H^T Y $$
Otherwise, the optimal output weight matrix is calculated according to equation (7).

\[
\hat{\beta} = H^T \left( \frac{1}{C} + HH^T \right)^{-1} Y
\]  
(7)

In equation (7),

\[
Y = \begin{bmatrix}
y_1^T \\
\vdots \\
y_N^T 
\end{bmatrix}_{N \times M}
\]

The calculation process is explained as follows [9, 10].

1. The theory has proved that as long as the number of single hidden layer nodes L is sufficient, the activation function \( G(a_i, b_i, x) \) can achieve infinitely differentiable in any interval, and no adjustment of network parameters is required under this premise.

2. The optimal output weight matrix has the following characteristics:
   
   (a) The estimation of output weight matrix \( \beta \) calculated by the least square method can make the ELM algorithm obtain the minimum training error.
   
   (b) The estimated value of \( \beta \) is the minimum normal form, which indicates that the ELM network has the best generalization ability.
   
   (c) The \( \hat{\beta} \) estimate is unique, so the ELM output is the global optimal solution rather than the local optimal solution.

The advantage of ELM model algorithm is that in the application, only the number of hidden layer nodes is set and the activation function is selected, without human interference, so the learning speed is fast, the generalization ability is good, and the calculation result obtained is the global optimal solution. Based on the above unique advantages, this paper attempts to use the ELM algorithm to identify the energy spectrum characteristics of cutting tool cutting sound signals to detect the working state of the running cutting tool.

4. Experimental Simulation and Result Analysis

4.1. Preparation for Experimental Simulation

The installation of the environment and sound sensor of a horizontal drilling machine on the test site is shown in figure 4. The time domain characteristics of drilling processing at a certain time when the tool is idling, initial normal, severe wear and tool collapse are shown in figure 5.

![Sound sensor](Image)
Figure 5. Time domain characteristics under different working states.

According to MATLAB spectrum analysis, when the tool is idling, the noise spectrum is distributed in the range of 300Hz ~ 1.6KHz. Under normal operation, the sound spectrum is in the range of 150Hz ~ 1000Hz. In the case of severe wear, the sound spectrum is distributed in two frequency bands, 300Hz ~ 1.4kHz and 1.8kHz ~ 3kHz. When the tool is in normal, serious wear and damage (broken edge) state, different cutting sound signal in the total energy of the energy of each band is different. Figure 6 records the proportion of energy of each frequency band in the total energy under the wavelet packet decomposition of five layers. In figure 6, each node on the horizontal axis is marked with three histograms, from left to right, showing the energy ratio of normal, severely worn, and damaged states respectively.

Figure 6. Proportion of energy in each frequency band.

In terms of the analysis of figure 6, the energy proportion of each node after node 17 is less than 0.01%, so it can be negligible. Based on this, the node of ELM input layer can be determined. According to the ratio of frequency band energy of nodes decomposed by 5-layer wavelet packet in figure 1, the feature input vector is 16, so the number of nodes in the input layer of the ELM model is selected as 16. The feature input vector is arranged in an array according to node order, and the working state of the cutting tool is represented in the form of energy spectrum feature vector.

Determine the number of nodes in the hidden layer. Obviously, if the number of nodes is too small, underfitting will occur and the identification accuracy will be reduced. On the contrary, overfitting will occur, resulting in long learning time and difficulty in identifying new samples. Based on the
combination of empirical formula, experiment and calculation, the repeated experiments show that when the hidden layer node is 118, the model recognition accuracy is the highest and the sample training time is the shortest. Therefore, the number of hidden layer nodes is determined to be 118.

The frequency band energy ratio of each node decomposition according to the five-layer wavelet packet is consistent with the energy spectrum features of the cutting sound signal. The sample data are collected based on MATLAB platform. In a large number of collected experimental data samples, 200 groups of data in each of the three states of normal cutter label No. 1, broken blade label No. 2 and severe wear label No. 3 were randomly selected, and a total of 600 groups of experimental data were taken as samples. The experimental data records of feature samples under different states were shown in table 1.

| Number | Node1 | Node2 | Node3 | ... | Node16 | Tool status       | Output label |
|-------|-------|-------|-------|------|--------|-------------------|--------------|
| 1     | 26.617| 33.074| 4.5827| ...  | 2.0145 | Normal cutter 1   | 1            |
| 2     | 27.069| 32.599| 4.4367| ...  | 2.0587 | Normal cutter 1   | 1            |
|       |       |       |       |      |        |                   |              |
| 200   | 30.246| 38.311| 5.3623| ...  | 1.0342 | Normal cutter 1   | 1            |
| 201   | 36.143| 27.006| 3.8861| ...  | 1.7925 | Broken blade 2    | 2            |
| 202   | 32.603| 31.646| 3.7744| ...  | 1.6903 | Broken blade 2    | 2            |
|       |       |       |       |      |        |                   |              |
| 400   | 41.741| 16.763| 4.2514| ...  | 1.8108 | Broken blade 2    | 2            |
| 401   | 11.201| 14.619| 15.285| ...  | 2.8076 | Severe wear 3     | 3            |
| 402   | 10.684| 13.027| 15.859| ...  | 3.5393 | Severe wear 3     | 3            |
|       |       |       |       |      |        |                   |              |
| 600   | 9.930| 14.915| 16.693| ...  | 2.1731 | Severe wear 3     | 3            |

4.2. Experimental Results and Analysis

In the experiment, the Sigmoid function was selected as the activation function of the ELM model, and all the energy spectrum feature sample data recorded in the energy spectrum feature sample in table 1 were normalized before the experimental simulation. First of all, the first 100 sets of data corresponding to the energy spectrum feature sample records in the three states of normal, severe wear and damage (broken edge) were selected as the learning and training samples of the ELM model after normalization treatment, and the parameter training of the ELM model was completed based on this. Then, the last 100 groups of data under the three conditions of normal, severe wear and damage (broken edge) were selected as test samples and then tested, which were used as the final test results of the experimental simulation. The final identification results of the tool working state are shown in table 2.

| State     | Status classification number | Test samples number | Correct identification number |
|-----------|------------------------------|--------------------|------------------------------|
| Normal    | 1                            | 100                | 93                           |
| Damage    | 2                            | 100                | 89                           |
| Severe wear | 3                          | 100                | 95                           |
As can be seen from table 2, the correct identification rates of its normal working state, damaged (broken blade) state and severe wear state are 93%, 89% and 95% respectively. Because the ELM model operates in the “on-line state” during the actual operation of cutting tool wear, its real-time performance is quite high in the practical engineering application.

5. Conclusions
It is very complicated to judge the fault of cutting tool wear during machining. Due to the limited experimental conditions such as machining site and data collection of tool type, the fault classification is not so detailed enough, and this paper has only completed the principle verification, so the research work is a little rough. There are still many problems to be further studied and solved before the method discussed can be put into practical production. However, the preliminary research results show that the sound recognition method based on ELM model is effective and applicable to detect the cutting tool wear state.

Acknowledgments
This work is supported by Chongqing Science and Technology Bureau Project of China (Grant no. cstc2019jscx-fxydX0090), and Key Project of Chongqing Science and Technology Bureau of China (Grant no. CSTC2019JSCX-FXYD0047).

References
[1] Zhang X C, Liu H Y, Shi Y C, et al. 2020, Experimental research on monitoring tool wear condition of vibration drilling based on acoustic emission signal Machine Tool & Hydraulics 48 (13) 189-192.
[2] Ai C S, He G W, Dong Q C, et al. 2009 Tool wear monitoring based on cutting sound LPCC China Mechanical Engineering 20 (17) 2045-2048.
[3] Ai C S, Dong Q C, Ze X B, et al. 2007 Tool breakage monitoring based on wavelet analysis of acoustic signal Modular Machine Tool & Automatic Manufacturing Technique (10) 54-56.
[4] Huang G B, Zhu Q Y and Siew C K 2006 Extreme learning machine: theory and applications Neurocomputing 70 (1) 489-501.
[5] Liu X 2012 Research on Weak Signal Extraction in Strong Noise Environment Based on Wavelet Analysis (Nanjing: Nanjing University) Master Dissertation pp 30-40.
[6] Mannan M A, Ashraf A, Kassim, et al. 2000 Application of image and sound analysis techniques to monitor the condition of cutting tools Pattern Recognition Letters (21) 969-979.
[7] Li T, Huang X Y and Luo M 2020 Analysis of tool wear characteristics based on wavelet packet decomposition Modular Machine Tool & Automatic Manufacturing Technique (07) 10-15.
[8] Song Q J 2016 Research and Application of ELM Neural Network (Nanjing: Nanjing University of Information Science & Technology) Master Dissertation pp 26-37.
[9] Guan S and Nie P 2012 Review and prospect of research on on-line metal cutting tool wear monitoring III: Pattern recognition method Machine Tool & Hydraulics 40 (03) 148-153.
[10] Luo D 2018 Research on Process Monitoring Method Based on Extreme Learning Machine (Guangzhou: South China University of Technology) Master Dissertation pp 41-52.