# Research of Tool Wear Based on Radical Basic Function Network

XU Chuangwen FAN Xing LUO Wencui

*Abstract*—A neural network simulator built for prediction of milling tool wear in cutting process is discussed. Experiments are conducted on an existing laboratory equipment to generate training and test data. The cutting current of the singular value decomposition (SVD) noise reduction and cutting parameters are employed to train network. Network train is carried out for a variety of inputs. The adaptability of different cutting conditions is investigated. The networks are validated for test data with unknown tool wear and established tool wear net model can calculate tool wear value in different cutting conditions satisfactorily.

*Keywords* — radical basic function network; tool wear; singular value decomposition.

#### I. INTRODUCTION

Wear and wear condition of metal cutting tool directly effect the precision, efficiency and economic benefit of machining process, so monitoring wear of cutting tool are becoming more and more significant. On-line tool wear monitoring is an important topic to flexible manufacture system. In recent years, several monitoring methods have been developed by researchers. But few can be applicable to practical production. Although some of them can be applied, it is difficult to popularize for practical application because of rigorous requirements and limitations to the working condition<sup>[1-12]</sup>.

In the cutting process, techniques for tool wear monitoring are being used widely using the current of machine tool spindle and feed<sup>[13-15]</sup>. It does not interference with cutting process by measurement equipment and the machine tool didn't formed by reworking process. But it is a key problem how to find mathematical models between cutting current and cutting parameters in research and develop monitor current.

In this paper, a tool wear model of radial basic function network(RBFN) is obtained based on cutting current .Tool wear net model is trained by selecting reasonable factors related to tool wear and adopting several groups' data of tool wear in different cutting conditions, then forecasts tool wear.

Manuscript received October 9, 2008.

#### II. RBFN

A radial basis function network (RBFN) is a hybrid learning neural network, which is a two-layer fully-connected network with an input layer which performs no computation. RBFN is a type of neural network using local receptive field. Here is the RBFN shape. Its structure is shown in Fig.1. It consists of following three layers:



## Fig.1 Structure of RBFN

1) Input t layer:  $X = (x_1, x_2 \cdots x_N)^T$  is a *N* dimension input vector.

2) The hidden layer: Learning in the hidden layer is performed by using an unsupervised method, the K-means algorithm. First, the user must choose a number of centers and this number will correspond to the number of neurons in the hidden layer. The K-means algorithm is used to position the centers in the best way, i.e. each presented record is attached to its nearest center (or cluster). As it is an unsupervised learning method, only the inputs data are presented to the K-means algorithm. The radial basis functions are then applied to each center: There is one radial Gaussian function for each hidden unit which simulates the effect of overlapping and locally tuned receptive fields. Weights between nodes in the input layer and in the hidden layer are identically equal to 1. The activation function of the hidden nodes is radially symmetric in input space; the magnitude of the activation given by a particular record is a decreasing function of the distance between the input vector of the record and the center of the basis function. The distance metric employed is the Euclidean distance. The activation of the hidden neuron *j* with presented pattern X has the form

$$u_{j} = \exp\left(-\frac{\|X - C_{j}\|^{2}}{2\sigma_{j}^{2}}\right) \qquad j = 1, 2, \cdots, m$$
 (1)

XU Chuangwen is with the Department of Mechanical Engineering, Lanzhou Polytechnic College, China (e-mail: xuchuangwen@ 126.com).

FAN Xing is with with the Department of Mechanical Engineering, Lanzhou Polytechnic College, China.

LUO Wencui is with with the Department of Mechanical Engineering, Lanzhou Polytechnic College, China .

Where X is the network input vector,  $X = (x_1, x_2 \cdots x_N)^T$ ;  $C_j = (c_{1j}, c_{2j} \cdots c_{Nj})^T$  is a vector with the same dimension as X and it is the center of the Gaussian function for *j* th hidden node;  $\sigma_j$  is the width of the central function, which is the range of the control function, also named variance. Where the Euclidean distance is

$$L_{j} = \left\| X - C_{j} \right\| = \sqrt{\sum_{i=1}^{N} (x_{i} - c_{ij})^{2}} \qquad j = 1, 2, \cdots, m \text{ ; and}$$
$$i = 1, 2, \cdots, N \qquad (2)$$

Each hidden unit has an associated  $\sigma$  'width' value which defines the nature and scope of the unit's receptive field response. It is equivalent to the standard deviation of the width of the Gaussian response, so larger values allow more points to be included.

The closer presenting pattern to the center, the higher the response. This response will always be between 0 and 1 (The response will be 1 only if the presented pattern is identical to the center). The width of the 'bell' shape depends on the value of  $\sigma$ : The bigger the value of  $\sigma$ , the larger the width of the 'bell' shape. The value of  $\sigma$  is computed according to the distance between the corresponding cluster and its 2 closest one. The user can't neither choose the value of  $\sigma$  nor the way it's computed.

The role of the hidden units is to perform a non-linear transformation of the input space into the space of activations of the hidden units. It is this representation that gives the RBF a much greater representational power than the linear perceptron.

3)The output layer: Learning in the output layer is performed by computing a linear combination of the activation of the basis functions, parameterised by the weights  $\omega$  between hidden and output layers. Objective function of net is as follows

$$e = \frac{1}{2N} \sum_{i=1}^{N} [\hat{y}(x_i) - y(x_i)]^2$$
(3)

Where *e* is the total average error;  $\hat{y}(x_i)$  is the actual output relative to input  $x_i$ ;  $y(x_i)$  is expectation output relative to input  $x_i$ ; *N* is numbers of training samples.

In order to make objection function reach the minimum, the gradient descent method can be used to modify weights  $\omega$  which is from net hidden layer to output layer:

$$\Delta \omega_j = -\eta \frac{\partial e}{\partial \hat{y}(x_i)} \cdot \frac{\partial \hat{y}(x_i)}{\partial \omega_j} \tag{4}$$

Where  $\mu$  is learning rate , its range of value is from 0 to 1. The adjusting amount of each step  $\omega$  is confirmed according to Equation (3) and (4) listed, and it is as follows

$$\Delta \omega_j = -\frac{\eta}{N} \sum_{i=1}^N \left[ \hat{y}(x_i) - y(x_i) \right] \mu_j \tag{5}$$

$$\omega_j = \omega_j + \Delta \omega_j \tag{6}$$

Net output is as follows

$$y(x) = \sum_{j=1}^{m} \omega_j u_j(x) + \theta \tag{(}$$

#### III. BACKGROUND THEORY

Measurable tool wear signals typically have very low signal-to-noise ratio because of the variety of noise sources in the milling process. However, relatively little work has been done on tool wear signal enhancement and noise reduction. For tool wear classification, most monitoring systems either use the noisy signals directly without pre-processing, or simply lowpass filter the signal to average out the corrupting noise sources While relatively easy to implement, these techniques have proven to be generally ineffective at reducing the noise and tend to remove information necessary for proper tool wear classification. To choose more effective signal enhancement and noise reduction algorithms, we first need to examine the tool wear signal and noise generating processes .The noise can exist in the obtained data because there is much error during the process of collecting the initial data and much system error of the machine tools itself as well. The noise reduction must be done before obtaining the sample data.

### A. The singular value decomposition (SVD)

In the singular value theory, SVD of  $m \times n$  dimension matrix X ( $r \leq r_{max} \leq \min(m, n)$ )that any of their rank is equal to r is shown as  $X = U\Lambda V^{T}$ , where U and V respectively is  $m \times r_{max}$  and  $n \times r_{max}$  dimensions orthogonal matrix, and  $UU^{T} = I$ ,  $VV^{T} = I$ .  $\Lambda = \operatorname{diag}(\lambda_{1}, \lambda_{2}, \dots, \lambda_{r \max})$  is a diagonal matrix and its main diagonal element is all nonnegative value and arranges according to following order:  $\lambda_{1} \geq \lambda_{2} \geq \dots \geq \lambda_{r \max} > 0$ , where  $\lambda_{1}, \lambda_{2}, \dots, \lambda_{r \max}$  is the singular value of the matrix X, U and V respectively is the left and right singular matrix of X.

Supposed the time series measured cutting current signal is x(i),  $i = 1, 2, 3, \dots, N$ , a  $m \times n$  dimensions matrix is composed according to certain method.

$$\boldsymbol{D}_{m} = \begin{bmatrix} x(1) & x(2) & \cdots & x(m) \\ x(2) & x(3) & \cdots & x(m+1) \\ \vdots & \vdots & \vdots & \vdots \\ x(n) & x(n+1) & \cdots & x(n+m) \end{bmatrix}$$
(8)

Where  $D_m$  is known as the reconstruction attractor

orbital matrix and it can be expressed as  $D_m = D + W$ . Dis a  $m \times n$  dimension matrix which is composed by non-noise time series and W is also a  $m \times n$  dimensions matrix which is composed by noise time series. SVD of  $D_m$  is as follows

$$\boldsymbol{D}_{\boldsymbol{m}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{U}(\boldsymbol{\Lambda}_{1} + \boldsymbol{\Lambda}_{2})\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{U}\boldsymbol{\Lambda}_{1}\boldsymbol{V}^{\mathrm{T}} + \boldsymbol{U}\boldsymbol{\Lambda}_{2}\boldsymbol{V}^{\mathrm{T}} = \hat{\boldsymbol{D}} + \hat{\boldsymbol{W}}$$
(9)

Where  $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_k, \dots, \lambda_{r \max})$  is a singular value matrix  $\Lambda_1 = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_k, 0, \dots, 0)$ and,  $\Lambda_2 = \operatorname{diag}(0, 0, \dots, 0, \lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_{r \max})$ . If the series signals are not of noise or its signal-to-noise ratio is very high,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k, 0, \dots, 0)$ ; if the series signals include noise or are of the low signal-to-noise ratio,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_{r \max})$ . In the calculating process, if the *m* and *n* is not too small, the attractor orbital matrix  $D_m$  must be singular and the noise signals *W* is a column with full rank matrix .According to SVD theory, *k* maximum singular value ( $k < r_{\max}$ ) is selected and other singular values are equal to 0, then a matrix  $\hat{D}$  is obtained in the reverse process of the singular value decomposition. Therefore, the slick response signals  $\hat{D}$  is an optimum approximation matrix without noise and its rank is *r*.

### B. Reconstruction method

The contribution rate of the singular value is as follows

$$s_i = \lambda_i^2 / \sum_{j=1}^{r_{\text{max}}} \lambda_j^2 \qquad j = 1, 2, \cdots, r_{\text{max}}$$
 (10)

If the signal is slick and don't include the noise, the first several values of  $\{s_i\}$  are big, and others are equal to 0. If the signal includes the noise,  $\{s_i\}$  is not equal to 0. Because the energy of the noise signal is wide on the frequency domain distribution, singular value is far less than the primary signals. Therefore, the value after the point in the  $s_i \sim i$  curve that the drops and then becomes smooth contributes to the noise. Make their singular value equal to 0, and a new matrix can be obtained.

#### C The Time Series

A method to reconstruct the phase space was proposed by Packard based on single variable time series<sup>[16]</sup>. Supposed the time series of a dimension observation is  $x_i = x(t_i)$ ,  $t_i = t_0 + i\Delta t$ ,  $i = 1, 2, \dots, N$ . *m* dimension phase space is constructed in terms of sampling with equal space length and time delay  $\tau$ ,  $\tau$  is integral times of  $\Delta t$ . *m* dimension phase space is defined as follows

$$X_{i}(m,\tau) = \{x_{i}, x_{i+\tau}, x_{i+2\tau}, \cdots, x_{i+(m-1)\tau}\}$$
  
$$i = 1, 2, \cdots, N - (m-1)\tau$$
(11)

Where *m* is embedding dimension,  $\tau$  is time delay and  $\tau = K\Delta t$ ,  $\Delta t$  is interval time between sampling data and K is random integer. According to Tankens' embedding theory, the method obtaining condition vector  $X_i$  from time series  $x_i$  is called time delay embedding method. Embedding dimension m and time delay  $\tau$  must be selected carefully in order to give really expression to the dynamical characteristic from the measuring signal based on time delay embedding method . Tankens' embedding theory fails to show the principle of selecting the time delay, but only consider that as long as embedding dimension fills with  $m \ge 2D + 1$ , reconstruction phase space and the system phase space are differential coefficient homeomorphism, that is, topology equivalence, their dynamical characteristic is completely similar in the qualitative sense. When Ddimension attractor can embed in  $m \ge 2D + 1$  dimension phase space, the geometry characteristic of the initial attractor can be reappeared, and the evolvement law of the system can be researched.

When the phase space is reconstructed, the selection of time delay  $\tau$  must assure that every component is relative independence. That is, the relativity of the phase space ordinate is as less as possible. Autocorrelation correlation function and mutual information method are very ordinary methods in selecting the time delay. In this thesis mutual information method is used to select the time delay because it is more advanced<sup>[17-18]</sup>.

Mutual information principle: supposed the states of the discrete variable X and Y are m and n, their entropy function is defined as follows

$$H(X) = -\sum_{i=1}^{m} p_i \ln p_i$$
(12)

Where  $p_i$  is probability which variable X appears in the *i* state. The combination entropy of the variable X and Y is defined

$$H(X,Y) = -\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \ln p_{ij}$$
(13)

Where  $p_{ij}$  is probability that variable X appears in the *i* state and variable Y appears in the *j* state. According to the definition of the entropy of X and Y and combination entropy of X and Y, the mutual information can be derived as follows

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$
(14)

The total dependency of two variables can be measured by the mutual information function. Because the mutual information value of the first minimum is less and the two-double inception is differentiated more clearly, the dynamic characteristic of the attractor can be analyzed qualitatively and qualitatively through reconstructing the phase space. It is a better method to select time delay. Therefore, the optimal value is ascertained by using the average mutual information method, that is, selecting time delay when the mutual information function reaches the minimum firstly as time delay  $\tau$  reconstructing the phase space. The relation between the mutual information and the time delay  $\tau$  is shown in Fig.3. According to the mutual information method, that is,  $\tau = 1$ .

Embedding dimension *m* is usually obtained from time series phase space reconstruction according to formula  $m \ge 2D + 1$ . But such embedding dimension is not sure of minimum embedding dimension. Although much large embedding dimension can reconstruct the phase space such calculation easily increases other statistic complexity and is easily disturbed by outside noise. So it is necessary to search a minimum embedding dimension to reconstruct completely the phase space. Selection of common embedding dimension has system saturation method, false neighboring method and Liangyue Cao method and so on. The method of selecting the embedding dimension was proposed by Liangyue Cao in 1997<sup>[19]</sup>. The method defined two parameters  $E_1(m)$  and  $E_2(m)$ ,

among them, the minimum embedding dimension m was decided by  $E_1(m)$ , and pointed out when  $E_1(m)$  tends to be steadily in along with the evolution, the corresponds value m is the minimum embedding dimension. At the same time,  $E_2(m)$  can not be used to obtain the minimum embedding dimension, but it has a very good function, that is, it can be used to distinguishing random series or chaotic series from time series .It is random series if  $E_2(m)$  is equal to 1 or is near to 1 to any m. Therefore, to real chaotic series,  $E_2(m)$  can not be equal to 1 to any m. Generally,  $E_2(m)$  tends to 1 to a real chaotic series. Thus, it is a direct and simple method to decide whether the time series has fractal characteristic of the chaos series. In Fig.4, the minimum embedding dimension extracted by Cao method is nearly 10. Therefore, the singular value decomposition matrix is consisted of 10×1015 dimensions where  $N - (m-1)\tau = 1024 - (10-1) \times 1 = 1015$ .



(c) The tool wear acute stage

Fig.3 The relation curve between mutual information



# IV. EXPERIMENTAL RESULTS

We select  $v \land f \land a_p \land I_s \land I_f$  as the input parameters of RBFN. They are representative of cutting speed, feed speed, cutting depth, cutting current of spindle and feed current. The tool wear *VB* is nonlinear function of these parameters and is in the form

$$VB = F(v, f, a_p, I_s, I_f)$$
(15)

The parameters ( $v \, \cdot \, f \, \cdot \, a_p \, \cdot \, I_s \, \cdot \, I_f$ ) concerned with the cutting tool wear is as input patterns of RBFN. VB is as the output target. The essence of training process is approximation to the function, which is nonlinear mapping from original input patterns to target output VB in the multidimensional space. If the effect factors of tool wear are covered reasonably by original input patterns, RBFN model can correctly estimate and forecast the cutting tool wear VB based on training RBFN.

To a particular CNC machine tools, there is a strict correspondence between  $v' \sim f'$  and  $a'_p$  detected from the drive unit and actual  $v \sim f$  and  $a_p$ . If the curves of this correspondence relation is established, the actual  $v \sim f$  and  $a_p$  can be obtained easily. In data acquisition, A/D conversion chip is used for each analog signal, chose the XZCS6AND7 to start up multiple A/D conversion at the same time through logic circuit, and read conversion data from every highway was latched. At firstly, the data was processed by using above-mentioned theory of the singular value decomposition and the method of noise reduction.

The vector of the working condition recognition is constituted by the parameters  $v \\finstress f$ and  $\Delta I \\cdot \Delta I = (I_t - I_{t-1})/T$ , In the equation, T is sampling cycle according to experiences,  $I_t$  is current at time of t,  $I_{t-1}$  is current at time of t-1. Whether the working condition recognition on machine tool is rapid feed, work feed or stop is judged by the vector  $(v \\finstress f \Delta I)$ . The rules of working condition recognition are shown in Tab.1.

In actual process, no-load current is calculated by no-load current model according to every axial speed measured, so effective cutting current is calculated. But no-load current is changed by many factors such as machine parts temperatures, lubrication etc, so original no-load current needs update. In monitoring process, if the machine tool is no-load (in actual cutting process, every working feed begins at no-load), the detected current is no-load current.

Tab.1 The working condition recognition

	parameters			condition	
v	f	$\Delta I$	rapid feed	working feed	stop
	$f \ge f_{high}$	$\Delta I \leq \Delta I_{high}$	+		
	$f < f_{low}$	$\Delta I < \Delta I_{low}$		+	
$v > v_{low}$	$f_{high} \geq f > f_{low}$	$\Delta I \geq \Delta I_{high}$			+

The model  $I_{si}(v)$  (or  $I_{fi}(f)$ ) is corrected by this no-load current, then the no-load current model is changed with real-time tracking current variety.

In order to testify the rationality and practicability of the RBFN model, a lot of experiments were done. In the experiment process, the high speed steal tool and the  $45^{\#}$  steal workpiece material are used to do it. Higher hardless material is used for the cutting experiment in order to run tool wear up. Take the turn from little to more to cut according to the tool wear. On different wear stage, the current value is collected about 12 sets and every value is the average of 640 sets data. The milling experimental condition is shown in Tab.2.

Tab.2 Cutting experiment condition					
Cutting	Material	High-speed steel			
tool	Туре	End milling cutter			
1001	Diameter(mm)	14			
Equip	oment	XKA714			
Milling	method	Climb milling			
Workspac	e material	Thermal refining 45 steel			
Cutting spe	ed/(m/min)	8.792~21.98			
Feed speed	/(mm/min)	20~35			
Cutting d	epth/mm	2~5			

96 groups experimental data of the spindle cutting current and feed current are obtained in different cutting conditions as the tool value is equal to 0.05mm  $\times 0.1$ mm  $\times 0.15$ mm  $\times 0.2$ mm  $\times 0.25$ mm  $\times 0.3$ mm  $\times 0.35$ mm  $\times 0.4$ mm. As it is shown in Fig.7 and Fig.8, every curve is obtained in the same cutting condition.



Fig. 7 The spindle cutting current



RBFN is trained by sample set constituted of the 96 groups of current data and cutting parameters. With the increase in the number of neurons, the scope of network squared error is observed by selecting appropriate initial weights and deviation. Then trained network weights and deviation are kept to memory. The simulation result of RBFN model is shown in Fig.9.

From Fig.9, the linear correlation coefficient R between the network forecast and measured values is 0.92, and the result is a good degree of satisfaction. In order to verify the reliability of the network recognition, firstly, 16 groups of new testing samples are obtained randomly according to Tab.2. Secondly, calculation value VB' is calculated by trained artificial neural network model. Finally, Compared with actual tool wear value VB, the result is shown in Tab.3.

Compared with the experimental result, prediction accuracy of RBFN model falls within the scope of the request. When the tool wear is small, the relative error of model is larger, this can be that effective current caused by the smaller tool wear is small, the measurement results is easily influenced by many other factors. When the tool wear is larger, the relative error of model is smaller. Compare with other artificial neural networks, the advantages of RBFN are obviously by calculation of the relevant literature. Perhaps it was because the models in linear was given by previous these literatures basically, the nonlinear nature of factors associated with tool wear is difficult accurately. Another reason can be others modeling methods ignore some relatively minor factors because of the factors

affecting tool wear are more complex, so the forecast accuracy is not particularly desirable[20-21].



Fig. 9 Network model simulation result

#### I. CONCLUSION

There is a consanguineous relation between the current of machine tool spindle and feed and the cutting tool wear in cutting process. In this paper, the non-linear mapping of the original model of multidimensional space to the cutting tool wear is constructed using the radial basis function network. This method does not need to build a mathematical model of the cutting tool wear, and has better adaptability because the artificial neural network function has the self-study, associative memory and non-linear map characteristic and can effectively remove the influence of uncertainty factors in cutting process. The experiments show the model is applicable to monitor the milling tool wear in cutting process. It can be used to accurately control the normal or fierce wear, and further studies are needed to monitor the initial stages of the wear.

#### REFERENCES

- C.W. Xu, H.I.Chen, "A Research of tool wear recognizing based on wavelet packet pretreated and neural network," *Journal of System Design and Dynamics, The Japan Society of Mechanical Engineers*, 2007, 1 (4): 611- 612.
- [2] P. W. Prickett, C. Johns. "An overview of approaches to end milling tool monitoring," *Machine Tools and Manufacture*, 1999, 39:105-122.
- [3] E. Jantunen, "A summary of methods applied to tool condition monitoring in drilling," *Machine Tools and Manufacture*, 2002, 42:997-1010.
- [4] L. Wang, M.G. Mehrabi, "Hidden Markov model-based tool wear monitoring in turning," *Manufacturing Science and Engineering*, 2002, 124(3):651-658.
- [5] J. Kopac, S. Sali, "Tool wear monitoring during the turning process," *Materials Processing Technology*, 2001, 113:312-316.
- [6] A. Dragos, "Tool condition monitoring in broaching," *Wear*, 2003, 254:370-382.
- [7] H. L. Wang, H. Shao, "On-line tool breakage monitoring in turning," *Materials Processing Technology*, 2003, 139:237-242.
- [8] M.A.Mannan, A.Kassim, "Application of image and sound analysis techniques to monitor the condition of cutting tools," *Pattren Recognition Letters*, 2000, 21:969-979.

- [9] D. E. Dimla "The Correlation of vibration signal features to cutting tool wear in a metal turning operation," *Materials Processing Technology*, 2002, 19(10):705-713.
- [10] K. Jemielnlak, L. Kwlatkowski, P. Wrzosek, "Diagnosis of tool wear based on cutting forces and acoustic emission measures as inputs to a neural network," *Intelligent Manufacturing*, 1998,9(5): 447-455.
  [11] P. Srinivasa Pai, P. K. Ramakrishna Rao, "Acoustic emission alysis
- [11] P. Srinivasa Pai, P. K. Ramakrishna Rao, "Acoustic emission alysis for tool wear monitoring in face milling," *Production Research*, 2002,40(5): 1081-1093.
- [12] S. V. Kamarthi, S. R. T. Kumara, P. H. Cohen, "Flank wear estimation in turning through wavelet representation of acoustic emission signals," *Manufacturing Science and Engineering*, 2000,122(1): 12-19.
- [13] H. Shao, H. L. Wang, X. M. Zhao, "A cutting power model for tool wear monitoring in milling," *Machine Tools & Manufacture*, 2004,44: 1503-1509.
- [14] C. W. Xu, "Design and research for tool wear monitoring," Journal of Harbin Institute of Technology, 2003, 35(1): 76-80.
- [15] L. L Xiao, S. K. Tso, "Drill wear monitoring based on current signals," *Wear*, 1999, 231: 172-178.
- [16] N. H. Packard, "Geometry from a time series," Phys Rev, 1980, 45 :712-716.
- [17] S.B. Wang, X. Y. Li, Y. D. Cao, "Deformation measurements using mutual information : principle and applications," *Optics and Lasers in Engineering*, 2005, 43 (7): 895-903.
- [18] D.Erdogmus, "Minimax mutual information approach for independent component analysis," *Neural Computation*, 2004, 16 (6): 1235-1252.
- [19] Liangyue Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physical D*, 1997, 110 (1): 43-50.
- [20] J.M.Zheng, "On identification of tool wear by multiple features fusion based on artificial neural network," *Mechanical Science and Technology*,2002, 21 (1): 111-114.
- [21] M.Q.Zhu, "Sensor fusion via neural networks for estimation of tool wear," ACTA Aeronautical ET Astronautic Sinica,2002, 18 (1): 121-116.