Structural Engineering and Mechanics, Vol. 31, No. 1 (2009) 57-74
DOI: http://dx.doi.org/10.12989/sem.2009.31.1.057

Milling tool wear forecast based on the partial least-squares regression analysis

Xu Chuangwen†

School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shanxi, China
Department of Mechanical Engineering, Lanzhou Polytechnic College, Lanzhou, Gansu, China

Chen Hualing

School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shanxi, China

(Received October 25, 2006, Accepted December 2, 2008)

Abstract. Power signals resulting from spindle and feed motor, present a rich content of physical information, the appropriate analysis of which can lead to the clear identification of the nature of the tool wear. The partial least-squares regression (PLSR) method has been established as the tool wear analysis method for this purpose. Firstly, the results of the application of widely used techniques are given and their limitations of prior methods are delineated. Secondly, the application of PLSR is proposed. The singular value theory is used to noise reduction. According to grey relational degree analysis, sample variable is filtered as part sample variable and all sample variables as independent variables for modelling, and the tool wear is taken as dependent variable, thus PLSR model is built up through adapting to several experimental data of tool wear in different milling process. Finally, the prediction value of tool wear is compare with actual value, in order to test whether the model of the tool wear can adopt to new measuring data on the independent variable. In the new different cutting process, milling tool wear was predicted by the methods of PLSR and MLR (Multivariate Linear Regression) as well as BPNN (BP Neural Network) at the same time. Experimental results show that the methods can meet the needs of the engineering and PLSR is more suitable for monitoring tool wear.

Keywords: partial least-squares regression; singular value decomposition; tool wear; cutting experiment.

1. Introduction

Wear and wear condition of metal cutting tool directly effect the precision, efficiency and economic benefit of machining process. The tool wear value obtained from the NC machine on time is compensated in time, which not only improves the machining accuracy of NC machine, but also provides the possibility to manage the tool life on time and to optimize the cutting parameters automatically. So tool wear monitoring is becoming more and more significant. On-line tool wear monitoring is an important topic to flexible manufacture system. In past decade, the research of monitoring tool wear, especially, the occurrence, development and evolvement of the tool breakage,

† Ph.D., Corresponding author, E-mail: xuchuangwen@126.com

has been developed so much, and some significant conclusions are obtained (Kopac and Sali 2001, Dragos 2003, Wang and Shao 2003). But the methods to monitor the tool wear have being researched now. Some methods are applied to the certain specific aspect, the others are on the tested phase.

At present, the methods to obtain the cutting tool wear include direct method and indirect method. The former is usually to measure the cutting tool wear value directly by using the optimal sensor, such as CCD pick-up head because touching the tool shape can't be reached in the cutting process (Mannan and Kassim 2000). But It is very difficult to measure its value on-line accurately in the cutting process. The latter is to calculate the wear value by measuring the cutting vibration signals Dimla 2002) or acoustic emission (AE) signals (Srinivasa 2002, Kamarthi and Kumara 2000). The techniques are still difficult to utilize in the real cutting process because of the complexity of real-time power source signals, it is not easy to extract the feature information of tool wear from complex signals in time-domain,frequency-domain. In addition, many prior methods have been developed for monitoring tool wear by measuring spindle and feed motor power(current) and have proved that the tool wear is very sensitive to the change of the cutting power (Xu and Chen 2007, Shao and Wang 2004, Xu 2003, Ertunc and Loparo 2001). In the cutting process, techniques for tool wear monitoring are being used widely using the spindle and feed motor power. It does not interference with cutting process by measurement equipment and the machine tool didn't formed by reworking process. However, generation mechanisms of the milling tool wear is more complex and in the view of various factors that affect tool wear, it is difficult to build the exact practical analysis model. Therefore,it is necessary to use experiment data to ensure the analysis and model. In some general methods, an explicit model is built by using Multivariate Linear Regression analysis method (Chen 2004, Xu and Wang 2006) or an implicit model is built by using Neural Network Palanisamy and Rajendran 2007). MLR method for monitoring tool wear by measuring spindle and feed motor power is to establish a mathematical model between milling cutting parameters and the classification by fuzzy pattern using MLR analysis. Then tool wear model for spindle and feed motor power is established. Tool wear value is predicted by tool wear model. Tool wear model is adjusted using cutting parameters to make it have better dynamic, fuzzy, real-time characteristics.So it will be effective to be used in the nonlinear predictive control systems. NN method for monitoring tool wear by measuring spindle and feed motor power is to establish a Neural Network model which contain milling cutting parameters and cutting power. Then tool wear network model is trained by using several experimental data of tool wear in different cutting process, Tool wear value is predicted by Network model. Several problems exist with this methods, namely : (1) It is diffcult to establish a exact practical analysis model between milling cutting parameters and tool wear. (2) The model based on spindle and feed motor power is used to recogniz tool wear,it can also cause larger error in different cutting process by using MLR method because tool wear model coefficients are fixed, that is, low-precision and limiting the applications. (3) The results of prediction usually are unstable because it is diffcult to overcome multicollinearity of variables using MLR method. (4) NN is hard to give a reasonable interpretation at factors influencing tool wear model.

PLSR statistical analysis module performs model construction and prediction of property using the Partial Least Squares technique. It is based on linear transition from a large number of original descriptors to a small number of orthogonal factors (latent variables) providing the optimal linear model in terms of predictivity. The new method developed rapidly both in theory and application in recent years. PLSR is a method to model by using sample data. It vill combined with basic function of regression modeling, principal component analysis and canonical correlation analysis. It is superior to the traditional MLR model. It can model in independent variables existing multiply linearity correlates, that is, the method has capabilies of auto-select independent variables. When number of sample are smaller than number of variables, regression modeling is allowed, this is helpful for the need of application. PLSR contains all former independent variables in finally model. each of regression coefficients on independent variables will give easy reasonable interpretation in regression model. Through extracting component, it is better the linear correlation between extracting component and tool wear. When regression model is reverted, it can give a explicit expression,this provides an available reference for the qualitative studying between milling cutting parameters and tool wear. In this paper,milling tool wear is predicted by using PLSR analysis algorithms. In the new different cutting process, milling tool wear was predicted by the methods of PLSR and MLR as well as BPNN at the same time. Through the comparison and analysis of the data of the experiment, conclusion is made.

2. The PLSR essential theory

Provided that there are two types of variables $\boldsymbol{X}$ and Y, $\boldsymbol{X}$ includes p components $(x_1, x_2, \ldots, x_p)$ and $\boldsymbol{Y}$ includes q components $(y_1, y_2, \ldots, y_q)$. As a lot of practical surveying data are collected, independent variable $\boldsymbol{X}$ is

$$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p]_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

Where dependent variable $\boldsymbol{Y}$ is

$$\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_q]_{n \times q} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nq} \end{bmatrix}_{n \times q}$$

In order to obtain the statistical relationship of the variable $\boldsymbol{X}$ and Y, firstly, the principal component $\boldsymbol{t}_1$ is drawn from independent variable $\boldsymbol{X}$, $\boldsymbol{t}_1$ is the linear combination of $x_1, x_2, \ldots, x_p$. Then, the principal component $\boldsymbol{u}_1$ is drawn from the dependent variable Y, $\boldsymbol{u}_1$ is the linear combination of $y_1, y_2, \ldots, y_q$. The following conditions must be satisfied when PLSR method draws the two principal components.

1. The variation information in $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ is drawn as much as possible.
2. Correlation degree of $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ can be reached to the maximum

These require that $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ can represent the original data $\boldsymbol{X}$ and $\boldsymbol{Y}$ as possible, and $\boldsymbol{t}_1$ can explain $\boldsymbol{u}_1$ well. After drawing $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$, the regression of $\boldsymbol{X}$ to $\boldsymbol{t}_1$ and $\boldsymbol{Y}$ to $\boldsymbol{t}_1$ must be applied individually. If the accuracy of the regressive equation is satisfied, the arithmetic is concluded. Or else, the second principal component is drawn through using the remainder information that $\boldsymbol{X}$ is explained by $\boldsymbol{t}_1$ and $\boldsymbol{Y}$ is explained by $\boldsymbol{t}_1$. The regression and draw do not stop until the accuracy is obtained. If there are *m* principal components $t_1, t_2, \ldots, t_m$ at last, the regression $\boldsymbol{Y}_k$ to $t_1, t_2, \ldots, t_m$ can be applied by using PLSR method, then the regressive equation of $\boldsymbol{Y}_k$ to $x_1, x_2, \ldots, x_p$ can be obtained, where $k = 1, 2, \ldots, q$ (Wang 2005).

The main calculation process. Standardize the original data $\boldsymbol{X}$ and $\boldsymbol{Y}$, and the standardized data can be compared together whether they have different dimension and order of magnitude or not. The equation is as follows

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, p \tag{1}$$

Where x_{ij} is the value of the component x_j of the variable $\boldsymbol{X}_j$ in i sample, $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ is the average value of varible $\boldsymbol{X}_j$, $S_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}$ is the standard deviation of the varible $\boldsymbol{X}_j$ and $\tilde{x}_{ij}$ is the value after standardizing x_{ij}. The data matrix $\boldsymbol{E}_0 = (\boldsymbol{E}_{01}, \boldsymbol{E}_{02}, \ldots, \boldsymbol{E}_{0p})_{n \times p}$ can represent the standardized $\boldsymbol{X}$, accordingly, $\boldsymbol{F}_0 = (\boldsymbol{F}_{01}, \boldsymbol{F}_{02}, \ldots, \boldsymbol{F}_{0q})_{n \times q}$ can represent $\boldsymbol{Y}$.

Drawing the principal components. $\boldsymbol{t}_1$ is the first component of $\boldsymbol{E}_0$, that is $\boldsymbol{t}_1 = \boldsymbol{E}_0\boldsymbol{w}_1$, where $\boldsymbol{w}_1$ is defined the first axis of $\boldsymbol{E}_0$, and it is a unit vector, $\|\boldsymbol{w}_1\| = 1$. $\boldsymbol{u}_1$ is the first component of $\boldsymbol{F}_0$, that is $\boldsymbol{u}_1 = \boldsymbol{F}_0\boldsymbol{c}_1$, where $\boldsymbol{c}_1$ is defined the first axis of $\boldsymbol{F}_0$, and it is also a unit vector, $\|\boldsymbol{c}_1\| = 1$. To Solute the following optimized equation: $\lim_{\boldsymbol{w}_1, \boldsymbol{c}_1} \langle \boldsymbol{E}_0\boldsymbol{w}_1, \boldsymbol{F}_0\boldsymbol{c}_1 \rangle$, btained

$$\begin{cases} \boldsymbol{w}_1^T\boldsymbol{w}_1 = 1 \\ \boldsymbol{c}_1^T\boldsymbol{c}_1 = 1 \end{cases}$$

To make $\boldsymbol{\theta}_1 = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1$ and the Lagrange arithmetic is applied, there is

$$\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{w}_1 = \boldsymbol{\theta}_1^2\boldsymbol{w}_1 \tag{2}$$

$$\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1 = \boldsymbol{\theta}_1^2\boldsymbol{c}_1 \tag{3}$$

From the Eqs. (2) and (3), it is known that $\boldsymbol{w}_1$ is the unit eigenvector of the max eigenvalue on the matrix $\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0$ and $\boldsymbol{c}_1$ is the unit eigenvector of the max eigenvalue on the matrix $\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0$. After $\boldsymbol{w}_1$ and $\boldsymbol{c}_1$ are obtained, the component $\begin{cases} \boldsymbol{t}_1 = \boldsymbol{E}_0\boldsymbol{w}_1 \\ \boldsymbol{u}_1 = \boldsymbol{F}_0\boldsymbol{c}_1 \end{cases}$ is obtained. The regressive equations of $\boldsymbol{E}_0$ and $\boldsymbol{F}_0$ to $\boldsymbol{t}_1$ are solved

$$\boldsymbol{E}_0 = \boldsymbol{t}_1\boldsymbol{p}_1^{\mathrm{T}} + \boldsymbol{E}_1 \tag{4}$$

$$\boldsymbol{F}_0 = \boldsymbol{t}_1\boldsymbol{r}_1^{\mathrm{T}} + \boldsymbol{F}_1 \tag{5}$$

Where $\boldsymbol{E}_1$ and $\boldsymbol{F}_1$ are residual matrix of the two equations, $\boldsymbol{p}_1$ and $\boldsymbol{r}_1$ are the regressive coefficient vector, their expressions are

$$\boldsymbol{p}_1 = \frac{\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{t}_1}{\|\boldsymbol{t}_1\|^2}, \quad \boldsymbol{r}_1 = \frac{\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{t}_1}{\|\boldsymbol{t}_1\|^2}$$

According to $\boldsymbol{E}_1$ and $\boldsymbol{F}_1$, the regressive equations of $\boldsymbol{E}_1$ and $\boldsymbol{F}_1$ to $\boldsymbol{t}_2$ are as follows

$$\boldsymbol{E}_1 = \boldsymbol{t}_2\boldsymbol{p}_2^{\mathrm{T}} + \boldsymbol{E}_2 \tag{6}$$

$$\boldsymbol{F}_1 = \boldsymbol{t}_2\boldsymbol{r}_2^{\mathrm{T}} + \boldsymbol{F}_2 \tag{7}$$

Where $\boldsymbol{E}_2$ and $\boldsymbol{F}_2$ are residual matrix of the two equations, and $\boldsymbol{p}_2$ and $\boldsymbol{r}_2$ are the regressive coefficient vector, their expression are

$$\boldsymbol{p}_2 = \frac{\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{t}_2}{\|\boldsymbol{t}_2\|^2}, \quad \boldsymbol{r}_2 = \frac{\boldsymbol{F}_1^{\mathrm{T}}\boldsymbol{t}_2}{\|\boldsymbol{t}_2\|^2}$$

Up to step m, according to $\boldsymbol{E}_{m-1}$ and $\boldsymbol{F}_{m-1}$, the unit eigenvector $\boldsymbol{w}_m$ of the maximum eigenvalue on the matrix $\boldsymbol{E}_{m-1}^{\mathrm{T}}\boldsymbol{F}_{m-1}\boldsymbol{F}_{m-1}^{\mathrm{T}}\boldsymbol{E}_{m-1}$ and the unit eigenvector $\boldsymbol{c}_m$ of the maximum eigenvalue on the matrix $\boldsymbol{F}_{m-1}^{\mathrm{T}}\boldsymbol{E}_{m-1}\boldsymbol{E}_{m-1}^{\mathrm{T}}\boldsymbol{F}_{m-1}$ can be found.

$$\boldsymbol{E}_{m-1} = \boldsymbol{t}_m\boldsymbol{p}_m^{\mathrm{T}} + \boldsymbol{E}_m \tag{8}$$

$$\boldsymbol{F}_{m-1} = \boldsymbol{t}_m\boldsymbol{r}_m^{\mathrm{T}} + \boldsymbol{F}_m \tag{9}$$

To ascertain the number of variable $\boldsymbol{t}_m$ and $\boldsymbol{u}_m$. The number of variable $\boldsymbol{t}_m$, $\boldsymbol{u}_m$ usually are decided by the predicted residual error sum of squares (*PRESS*), that is, on every step, *PRESS* is residual square sum of estimated value and actual value on independent variable after a sample being lost.

$$PRESS_{(m)} = \sum_{k}^{l}\sum_{i}^{n}(y_{ik} - \tilde{y}_{km(-i)})^2 \tag{10}$$

If $PRESS_{(m)} - PRESS_{(m-1)}$ is lower than the prearranged accuracy, the iteration step ends, or else, $\boldsymbol{t}_m$ and $\boldsymbol{u}_m$ are drawn continuously to calculate iteratively (Tang 2002).

$$\boldsymbol{E}_0 = \boldsymbol{t}_1\boldsymbol{p}_1^{\mathrm{T}} + \boldsymbol{t}_2\boldsymbol{p}_2^{\mathrm{T}} + \ldots + \boldsymbol{t}_m\boldsymbol{p}_m^{\mathrm{T}} \tag{11}$$

$$\boldsymbol{F}_0 = \boldsymbol{t}_1\boldsymbol{r}_1^{\mathrm{T}} + \boldsymbol{t}_2\boldsymbol{r}_2^{\mathrm{T}} + \ldots + \boldsymbol{t}_m\boldsymbol{r}_m^{\mathrm{T}} \tag{12}$$

Because $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_m$ are all expressed by the linear combination of $\boldsymbol{E}_{01}, \boldsymbol{E}_{02}, \ldots, \boldsymbol{E}_{0p}$, the regressive equation of $\boldsymbol{Y}^* = \boldsymbol{F}_0$ to $\boldsymbol{x}_j^* = \boldsymbol{E}_{0j}$ can be obtained.

$$\boldsymbol{Y}^* = \boldsymbol{F}_{0k} = \boldsymbol{a}_{k1}\boldsymbol{E}_{01} + \boldsymbol{a}_{k2}\boldsymbol{E}_{02} + \ldots + \boldsymbol{a}_{kp}\boldsymbol{E}_{0p} + \boldsymbol{F}_k \tag{13}$$

Where $\boldsymbol{F}_k$ is residual matrix. According to the reverse process of standard, $\boldsymbol{Y}^*$ reverts to $\boldsymbol{Y}$.

3. The basis theory of the singular value

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 monitoring tool wear, most monitoring systems either use the noise 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. To choose more effective signal, the improved signal is proposed by using adaptive reduction algorithm. The noise can exist in the collect data because there is much error in collecting initial data and much system error of the machine tools itself as well. The noise reduction must be done before obtaining the sample data.

3.1 The singular value decomposition (SVD)

In the singular value theory, SVD of $m \times n$ dimension matrix $\boldsymbol{X}$ ($r \le r_{\max} \le \min(m,n)$) that any of their rank is equal to r is shown as $\boldsymbol{X} = \boldsymbol{U\Lambda V}^{\mathrm{T}}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively is $m \times r_{\max}$ and $n \times r_{\max}$ dimensions orthogonal matrix, and $\boldsymbol{UU}^{\mathrm{T}} = \boldsymbol{I}$, $\boldsymbol{VV}^{\mathrm{T}} = \boldsymbol{I}$. $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{r\max})$ is a diagonal matrix and its main diagonal element is all nonnegative value and arranges according to following order: $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{r\max} > 0$, where $\lambda_1, \lambda_2, \ldots, \lambda_{r\max}$ is the singular value of the matrix $\boldsymbol{U}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively is the left and right singular matrix of $\boldsymbol{X}$.

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

$$\boldsymbol{D}_m = \begin{bmatrix} x(1) & x(2) & \ldots & x(m) \\ x(2) & x(3) & \ldots & x(m+1) \\ \vdots & \vdots & \vdots & \vdots \\ x(n) & x(n+1) & \ldots & x(n+m) \end{bmatrix} \tag{14}$$

Where $\boldsymbol{D}_m$ is known as the reconstruction attractor orbital matrix and it can be expressed exactly as $\boldsymbol{D}_m = \boldsymbol{D} + \boldsymbol{W}$. $\boldsymbol{D}$ is a $m \times n$ dimension matrix which is composed by non-noise time series and $\boldsymbol{W}$ is also a $m \times n$ dimensions matrix which is composed by noise time series. SVD of $\boldsymbol{D}_m$ is as follows

$$\boldsymbol{D}_m = \boldsymbol{U\Lambda V}^{\mathrm{T}} = \boldsymbol{U}(\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)\boldsymbol{V}^T = \boldsymbol{U\Lambda}_1\boldsymbol{V}^T + \boldsymbol{U\Lambda}_2\boldsymbol{V}^T = \widehat{\boldsymbol{D}} + \widehat{\boldsymbol{W}} \tag{15}$$

Where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_{r\max})$ is a singular value matrix and $\boldsymbol{\Lambda}_1 = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k, 0, \ldots,$ $0)$, $\boldsymbol{\Lambda}_2 = \mathrm{diag}(0, 0, \ldots, 0, \lambda_{j+1}, \lambda_{j+2}, \ldots, \lambda_{r\max})$. If the series signals are not of noise or its signal-to-noise ratio is very high, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k, 0, \ldots, 0)$; if the series signals include noise or are of the low signal-to-noise ratio, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{j+1}, \lambda_{j+2}, 0, \ldots, 0)$. In the calculating process, if the m and n is not too small, the attractor orbital matrix $\boldsymbol{D}_m$ must be singular and the noise signals $\boldsymbol{W}$ 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 $\widehat{\boldsymbol{D}}$ is obtained in the reverse process of the singular value decomposition. Therefore, the slick response signals $\widehat{\boldsymbol{D}}$ is an optimum approximation matrix without noise and its rank is r.

3.2 Reconstruction method

The contribution rate of the singular value is as follows

$$s_i = \lambda_i^2 / \sum_{j=1}^{r_{\max}} \lambda_j^2 \qquad j = 1, 2, \ldots, r_{\max} \tag{16}$$

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.

3.3 The phase space reconstruction

A method to reconstruct the phase space was proposed by Packard based on single variable time series (Packard 1980). Supposed that the time series of a dimension observation is $x_i = x(t_i)$, $t_i = t_0 + i\Delta t$, $i = 1, 2, \ldots, N$. m dimension phase space is constructed in terms of sampling with equal space length and time delay τ, τ is integral times of Δt. m dimension phase space is defined as follows

$$X_i(m, \tau) = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, \tau_{i+(m-1)\tau}\} \qquad i = 1, 2, \ldots, N-(m-1)\tau \tag{17}$$

Where m is embedding dimension, τ is time delay and $\tau = K\Delta t$, Δt 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 τ 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 $m \geq 2D + 1$, 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 D dimension attractor can embed in $m \geq 2D + 1$ 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 τ 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 (Wang and Cao 2005, Erdogmus 2004). Mutual information principle: supposed the states of the discrete variable X and Y are m and n, their entropy function is defined as follows

$$H(X) = -\sum_{i=1}^{m} p_i \ln p_i \tag{18}$$

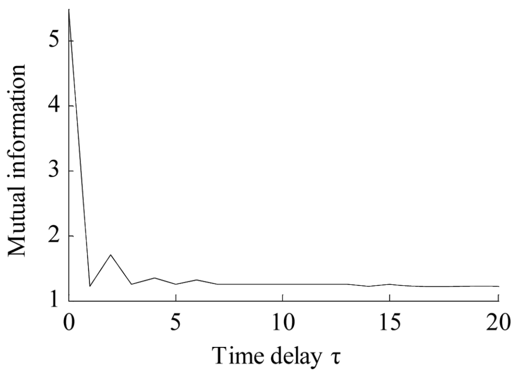
Where p_i is probability which variable X appears in the i state. The combination entropy of the variable X and Y is defined

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\ln p_{ij} \tag{19}$$

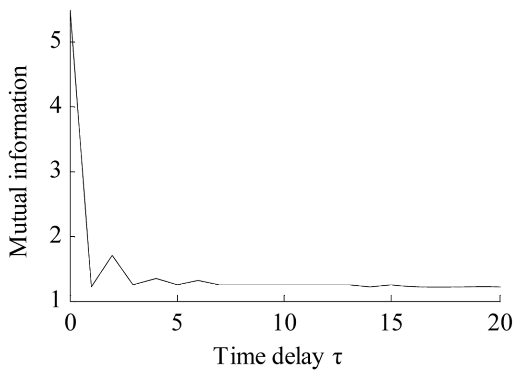
Where p_{ij} is probability that variable X appears in the i state and variable Y appears in the j state. According to the definition of the entropy of X and Y and combination entropy of X and Y, the mutual information can be derived as follows

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{20}$$

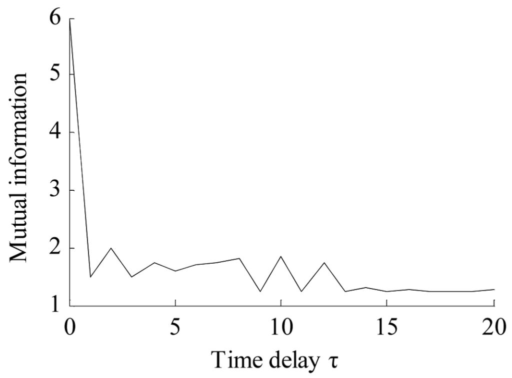
The total dependency of two variables can be measured by the mutual information function. Because the mutual information value of the first minimum is less and the two-double inception is differentiated more clearly, the dynamic characteristic of the attractor can be analyzed qualitatively and qualitatively 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 τ reconstructing the phase space. The relation between the mutual information and the time delay τ is shown in Fig. 1. According to the mutual information method, the time delay of the time series signals of the current is ascertained, that is, $\tau = 1$.



(a) The tool wear initial stage(*VB*=0.08)

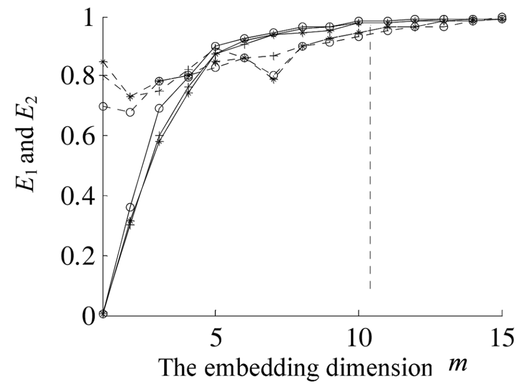


(b) The tool wear normal stage(*VB*=0.17)



(c) The tool wear acute stage (*VB*=0.30)

Fig. 1 The relation curve between mutual information



Full line indicates $E_1(m)$. Dotted line indicates $E_2(m)$. * indicates the tool wear initial stage
+ indicates the tool wear normal stage. o indicates the tool wear acute stage

Fig. 2 The minimum calculation dimension of Gao method

Embedding dimension m is usually obtained from time series phase space reconstruction according to formula $m \geq 2D+1$. 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. 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. 2, the minimum embedding dimension extracted by Cao method is nearly 10. Therefore, the singular value decomposition matrix is consisted of 10×1015 dimensions where $N-(m-1)\tau = 1024-(10-1)\times 1 = 1015$.

3.4 Revision and test

Firstly, the time delay τ and embedding dimension m is obtained by using the mutual information method and the minimum embedding dimension. Then the matrix $P(X_i)$ is composed of the time series x_i. $r_{\max}$ singular values of $P(X_i)$ are obtained by using the singular value decomposition. According to the singular value contribution factor s_i~i curve k maximum singular values $(k < r_{\max})$ are selected and other singular values is equal to 0. $P'(X_i)$ can be calculated and be transformed to $u_i, \ldots, u_{i+n-1}$ using the reverse process, where series $u_i, \ldots, u_{i+n-1}$ includes less noise than series x_i. In order to obtain the ideal signals, every $u_i, \ldots, u_{i+n-1}$ must be added correspondently and be averaged. Thus the signal x_i' after noise reduction is the ideal signal. Fig. 3 illustrates the time

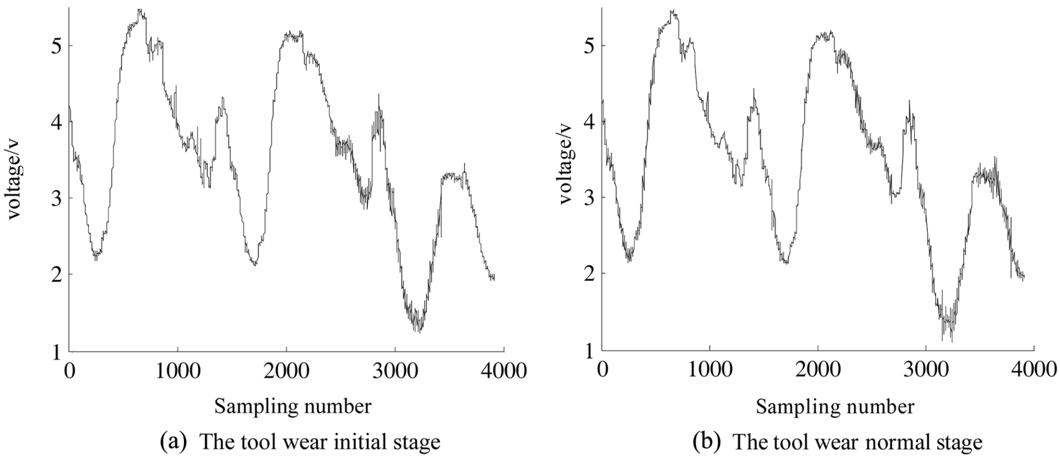


Fig. 3 The spindle power signals

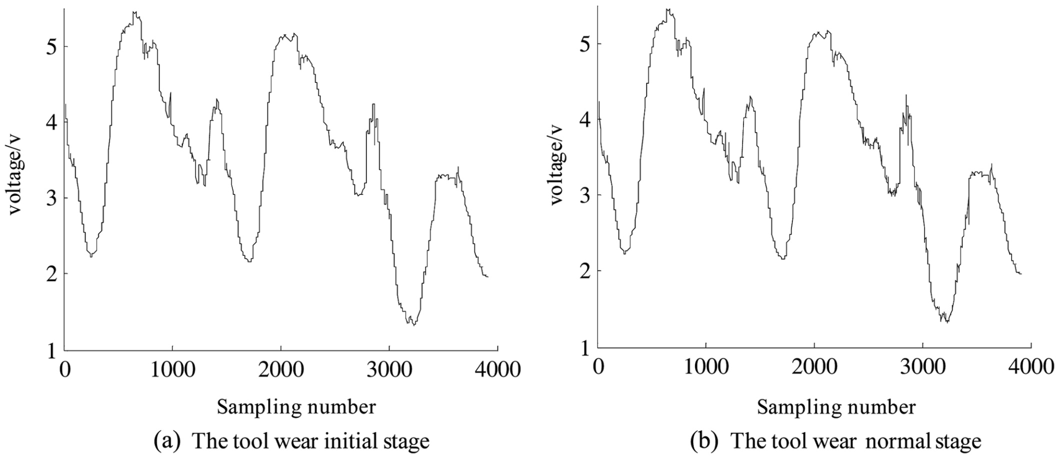


Fig. 4 The power signals of noise reduction

domain shape of spindle power signals in the tool wear initial stage ($VB = 0.08$) and the tool wear normal stage ($VB = 0.17$). In both the initial stage and the normal stage, the time domain shape distinction of power signals is not very obvious. The result shows that a lot of noise is contained in signals.In order to outstanding information characteristics of the tool wear, original data is reducted using noise reduction method of SVD, power signals of noise reduction are shown in Fig. 4.

4. The application

The monitoring system is composed of the measuring component, data collection, data processing and output monitor. The collection parameters are the input voltage, current of the spindle and feed motor. The collection instrument completes conversion of the signal and exchange of the data with the upper computer according to MODUBUS RTU agreement on time-sharing. The input modal vector consists of exchange of the data. The monitoring system of milling tool wear is shown in

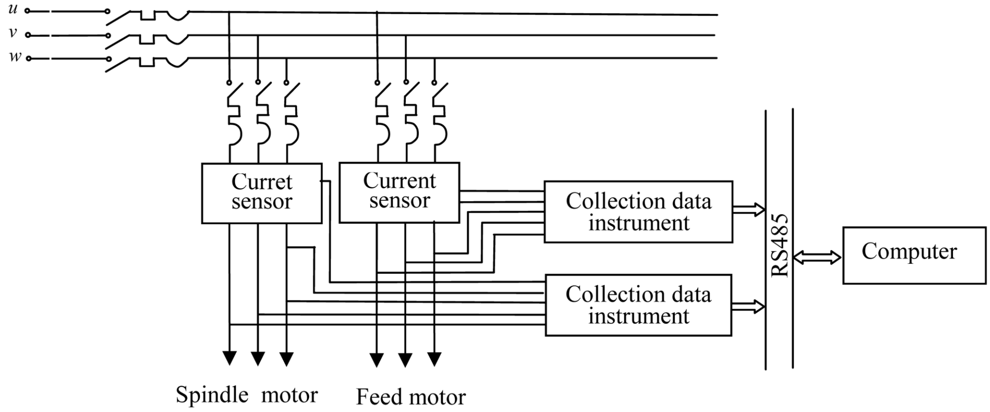


Fig. 5 The monitoring system of milling tool wear

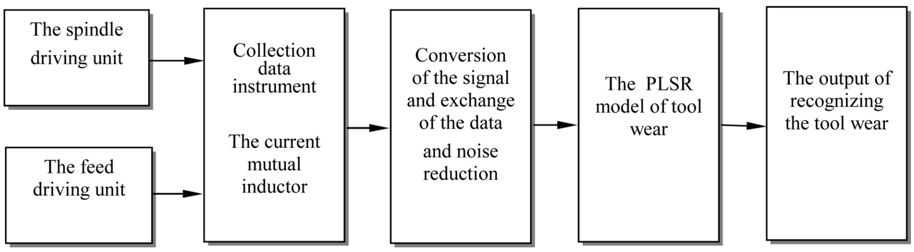


Fig. 6 The recognizing model of milling tool wear

Fig. 5. The recognizing model of milling tool wear is shown in Fig. 6.

NC milling XKA714 is used as experiment equipment in the cutting process. The experimental conditions include high-speed steel is as tool material, 45 steel is as workpiece material after being hardened and tempered, the hardness is HRC38~50, and the tool wear value *VB* is recorded with tool microscope. The sample data consists of the 96 sets of data which are collected on the different cutting conditions. Using the above-mentioned theory of SVD and the method of noise reduction, the noise were disposed from the power signals which data sample cycle is 1 second and the sample number is equal to 1024. According to the mutual information method, the time delay of the time series signals on the power is ascertained, that is, $\tau = 1$. According to Gao method, the minimum embedding dimension of the attractor orbit matrix is about to 10, thus the embedding dimension is equal to 10 in this research, or $m = 10$. Therefore, the SVD matrix is consisted of 10×1015 dimensions where $N-(m-1)\tau = 1024-(10-1)\times 1 = 1015$.

As it is shown in Fig. 7 and Fig. 8, every curve is obtained in the same cutting condition (the cutting speed is 8.792~21.98 m/min, the feed speed is 20~35 mm/ min, the cutting depth is 2~5 mm, the value of the tool wear respectively is 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, Table 1 is the example of 8 sets of the samples). v is cutting speed, f is feed speed, a_p is cutting depth, P_s is the spindle power, P_f is the feed power and *VB* is the tool wear value.

In the cutting process, the sample variable is screened firstly before the model is made because there are many factors influenced on the tool wear. The grey relational degree analysis method is applied on this thesis. Standardize the sample data, The method regards the tool wear serial as primary-serial $\{X_0(t)\}$ and the other as sub-serials $\{X_i(t)\}$, then formula for computing relational

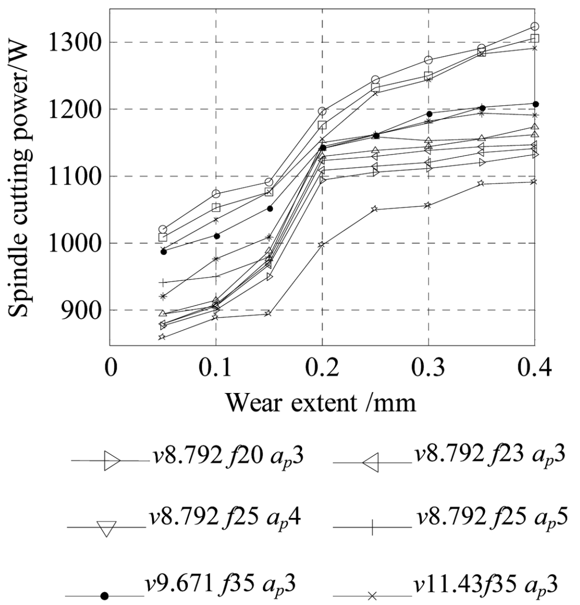


Fig. 7 The sample data of spindle power

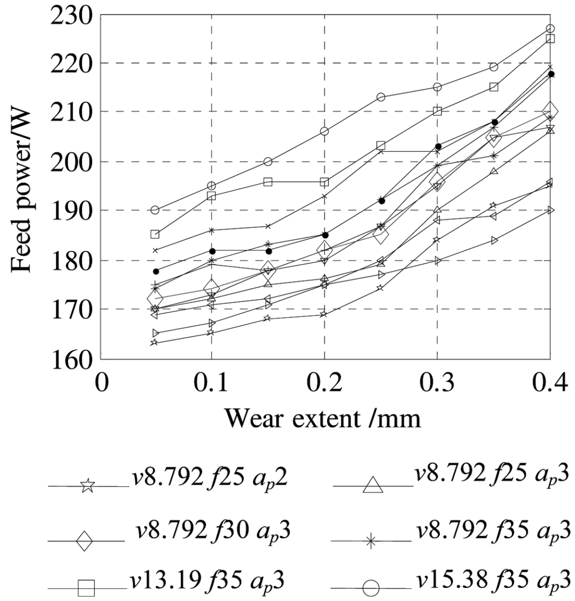


Fig. 8 The sample data of feed power

Table 1 Example of tool wear sample

No.	v/m·min^{-1}	f/mm·min^{-1}	a_p/mm	P_s/w	P_f/w	VB/mm
1	8.792	20	3	860	165	0.05
2	13.19	35	3	1036	186	0.1
3	8.792	25	5	980	183	0.15
4	11.43	35	3	1145	185	0.2
5	15.386	35	3	1233	203	0.25
6	8.792	20	3	1057	180	0.3
7	9.671	35	3	1195	207	0.35
8	11.43	35	3	1208	218	0.4

coefficient of primary-serial $\{X_0(t)\}$ and sub-serials $\{X_i(t)\}$ at time $t = k$ is as follows

$$L_{0i}(k) = \frac{\Delta_{\min} + \Delta_{\max}}{\Delta_{0i}(k) + \rho\Delta_{\max}} \tag{21}$$

Where $\Delta_{0i}(k)$ represents absoluteness difference of two compared serials at k time, namely $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$, $(1 \le i \le m-1)$. Δmax represents the maximum of absoluteness difference in all compared serials at every time. Δmin represents the minimum of absoluteness difference in all compared serials at every time. ρ is distinguishing coefficientits range of values is from 0.1 to 0.5.

$$r_{0i} = \frac{1}{N}\sum_{k=1}^{N} L_{0i}(k) \tag{22}$$

Eq. (22) is relational degree of primary-serial $\{X_0(t)\}$ and sub- serials $\{X_i(t)\}$, and the greater r_{0i} is, the closer serials $\{X_0(t)\}$ and $\{X_i(t)\}$ are. When $\{X_v(t)\}, \{X_f(t)\}, \{X_{a_p}(t)\}, \{X_{P_s}(t)\}$,

Table 2 Relational degree sort of sample variable series

Series	$\{X_v(t)\}$	$\{X_f(t)\}$	$\{X_{a_p}(t)\}$	$\{X_{P_s}(t)\}$	$\{X_{P_f}(t)\}$
Relational degree	0.6888	0.5981	0.7001	0.8119	0.8079
Sort	4	5	3	1	2

$\{X_{P_f}(t)\}$ and $\{X_{VB}(t)\}$ are taken into Eq. (21) and Eq. (22), Table 2 shows the relational degree of sample variable series.

From Table 2, it is found that the relational degree between the tool wear and feed speed are lower than other sample variables. It is very important to test and compensate the tool wear on-line if the effect of the feed speed in the modelling process by using PLSRA method is ignored because the feed speed can be adjusted in the actual cutting process in order to delete the effect of the variable of the feed speed on testing. Because the feed speed usually is adjusted on the actual machine process, the variation feed speed can't affect the test when it is ignored. According to this analysis, all sample variable (v, f, a_p, P_s, P_f) and part variable (v, a_p, P_s, P_f) are taken as independent variable and *VB* as the dependent variable. Thus the parameters of PLSR model can be obtained and researched. Then according to PLSR calculation step Eqs. (1)~(13) as well as the modeling sample data, and reference dropped trend of *PRESS* statistic and error statistic, the regressive model can be built when the latent variable is selected.

Fig. 9 is static drop trend on part sample error and Fig. 10 is static drop trend on all sample variable error. From Fig. 9 and Fig. 10, the error of the *PRESS* statistic and the error statistic are close to 0 when the latent variable is 3, thus the regressive model is set up by using them. After the relative parameter has been selected, the standard regressive coefficient of the dependent variable effected by the independent variable can be got as Table 3.

The regressive mathematic model of tool wear based on all sample variables and part variable can be obtained by using the reversal calculation on standard regressive coefficient according to Table 3.

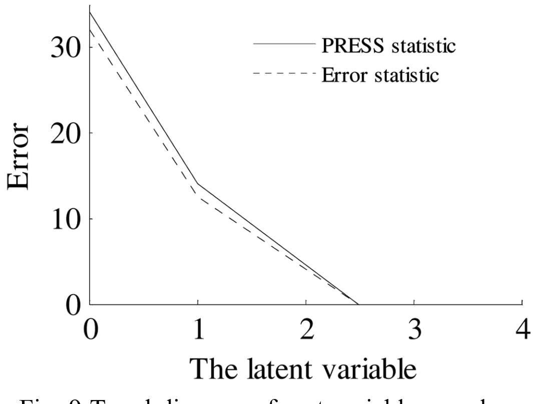


Fig. 9 Trend diagram of part variable sample error

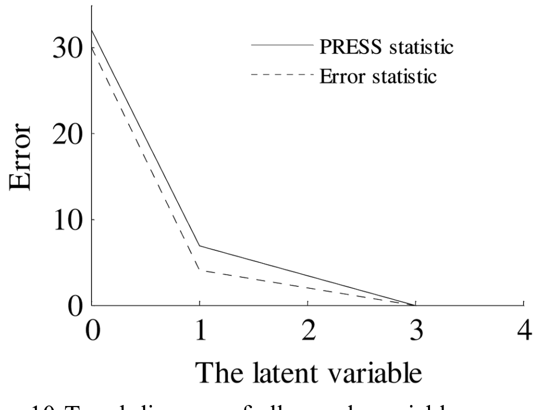


Fig. 10 Trend diagram of all sample variable error

Table 3 Standard regression coefficient of independent variable effected by dependent

Independent variable	v	f	a_p	P_s	P_f
Dependent variable VB_1 part	−0.5361	—	−0.1485	0.6031	0.5068
Dependent variable VB_2 all	−0.3369	−0.4287	−0.2142	0.4714	0.7414

Table 4 Error square sum and decision coefficient after data standardization

The num of the latent variable	Error square sum		Decision coefficient R^2		*PRESS* statistic	
	VB_1'	VB_2'	VB_1'	VB_2'	VB_1'	VB_2'
1	32.3420	32.3420	0.6596	0.6596	33.6155	34.1398
2	12.0234	5.4425	0.8734	0.9427	12.7070	5.9432
3	11.4870	3.4861	0.8791	0.9633	11.9831	3.7856

$$VB_1' = -0.758926 - 0.029223v - 0.024757a_p + 0.000586P_s + 0.003776P_f \tag{23}$$

$$VB_2' = -0.765035 - 018363v - 0.008942f - 0.03571a_p + 0.000458P_s + 0.005524P_f \tag{24}$$

From the static data of the Table 4, the fitting effect of the regressive model of the tool wear can be shown. It is also shown the drop of the sum of model error squares and the drop of *PRESS* statistic in the state of being standardizing the data, at the same time, the decision coefficient of the fitting the relative model is obtained, from which it can be seen that the fitting degree of the regression model is better when 3 latent variable are drawn out.

5. The verifying data and on-line forecast

5.1 The verifying of the model data

The model sample data was verified using Eq. (23) and Eq. (24) and tool wear VB' also can be obtained. Fig. 11 was the verify result of part sample variable, and Fig. 12 was the verify result of all sample variable. Many calculation value are close to the actual value from Fig. 11 and Fig. 12 as well as Table 5. By comparison with the tool wear, we are told, the fitting effect of all sample variables calculation value VB_1' are superior to the calculation value VB_1' of the part sample variable whenever the mean point of view or absolute error are taken into account. It shows that the feed speed and other cutting parameters are strongly correlated with the tool wear. The estimated value of the regressive coefficient will be changed a lot and will affect the value of the estimated tool wear, when the feed speed is ignored.

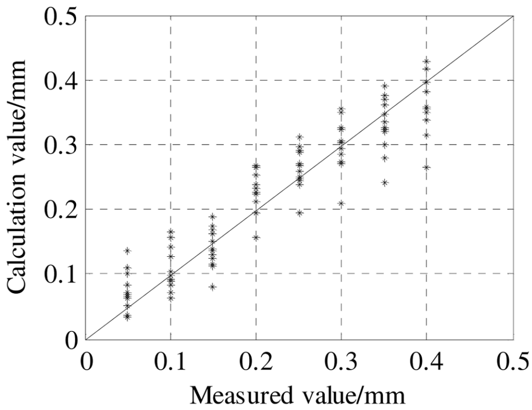


Fig. 11 Verify result of part sample variable

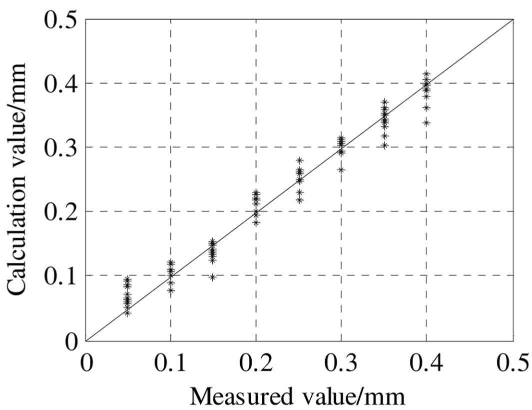


Fig. 12 Verify result of all sample variable

Table 5 The absolute difference of tool wear in different sample variable

Measuring value VB		0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
Calculation value of part sample variable VB_1'	Mean	0.074	0.106	0.140	0.226	0.263	0.299	0.331	0.364
	Absolute difference	0.085	0.064	0.07	0.066	0.061	0.091	0.108	0.134
Calculation value of all sample variable VB_2'	Mean	0.068	0.103	0.136	0.211	0.254	0.302	0.341	0.387
	Absolute difference	0.046	0.021	0.052	0.023	0.033	0.034	0.048	0.064

5.2 On-line forecast

In order to test whether the model of the tool wear can adopt to new measuring data on the independent variable, the spindle power and feed power are measured in the new different cutting process (the cutting speed is 6.594~26.376 m/min, the feed speed is 10~35 mm/min the cutting depth is 2~5 mm, the value of the tool wear respectively is 0.075, 0.125, 0.175, 0.225, 0.275, 0.325, 0.375, 0.425, Table 6 is the example of the 8 sets of sample). Thus, the tool wear that can be estimated on-line in the actual cutting process is forecasted. The forecast result of calculation value VB' compares with measuring value VB (Fig. 13 and Fig. 14). It can be seen from Fig. 13 and Fig. 14 that the better forecast value is obtained through calculating the mathematical model of the tool wear.

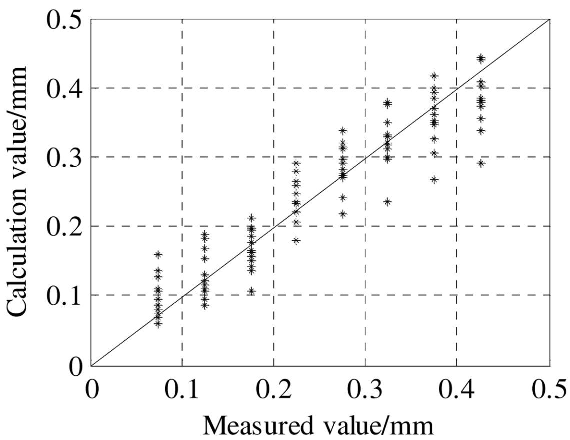


Fig. 13 Forecast result of part sample variable

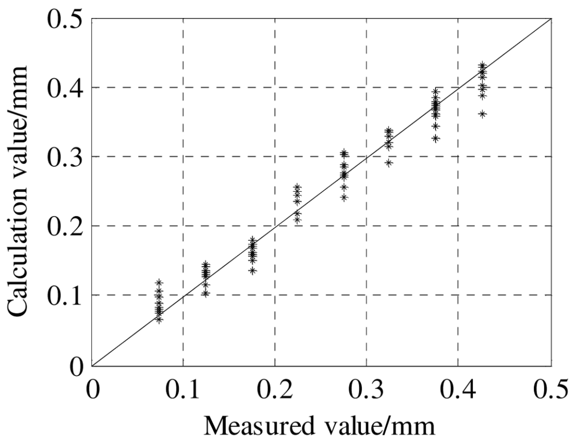


Fig. 14 Forecast result of all sample variable

Table 6 On-line forecast sample example of tool wear

No.	v/m·min^{-1}	f/mm·min^{-1}	a_p/mm	P_s/w	P_f/w	VB/mm
1	8.792	20	3	875	166	0.075
2	13.19	35	3	1031	189	0.125
3	9.671	35	3	1032	182	0.175
4	11.43	35	3	1116	190	0.225
5	15.386	35	3	1220	210	0.275
6	8.792	35	3	1171	193	0.325
7	9.671	35	3	1198	206	0.375
8	11.43	35	3	1286	214	0.425

5.3 Comparison with other methods

In the new different cutting process, milling tool wear was predicted by the methods of PLSR and MLR (Multivariate Linear Regression) as well as BPNN (BP Neural Network) at the same time. The results of the prediction and experiment were shown in Fig. 15 and Table 7. Through the comparison and analysis of the data of the experiment, to PLSR method, the maximal absolute error of the results of the prediction and experiment is within 0.035mm, the average error is within 0.028mm, which can meet the needs of the engineering; to MLR method, the maximal absolute error of the results of the prediction and experiment is within 0.053mm, the average error is within 0.047mm, which can meet the needs of the engineering. But we found the results of the prediction on the method are very unstable in a lot of the experiment. Problem may arise for different reasons, but multicollinearity of variables using MLR method can lead to main one. When linearly correlates of model are tested to be outstanding by F-test, t-test of almost all regression coefficients is not outstanding, the positive and the negative of regression coefficients are contrary results to expectation. The result of tool wear prediction turned out contrary to our expectations. If there is one or more to outstanding about the correlation coefficients, this shows that all independent

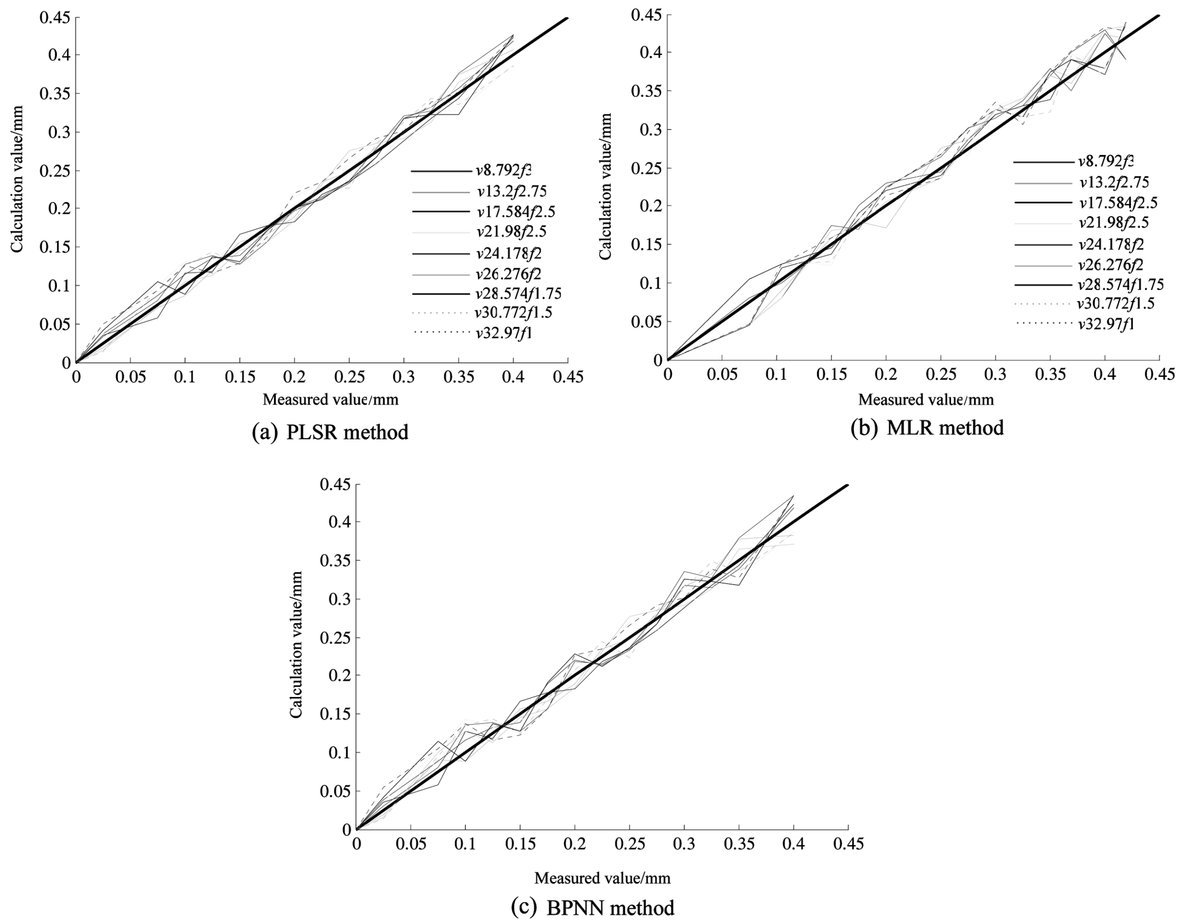


Fig. 15 Forecast results of three method

Table 7 Error statistics on tool wear (unit/mm)

Methods		Groups								
		1	2	3	4	5	6	7	8	9
PLSR	Maximal absolute error	0.029	0.028	0.027	0.032	0.031	0.030	0.035	0.035	0.030
	Average error	0.027	0.026	0.025	0.030	0.029	0.028	0.032	0.031	0.027
MLR	Maximal absolute error	0.050	0.053	0.049	0.051	0.047	0.047	0.050	0.048	0.048
	Average error	0.045	0.051	0.046	0.049	0.046	0.045	0.047	0.046	0.047
BPNN	Maximal absolute error	0.041	0.045	0.042	0.042	0.046	0.047	0.040	0.041	0.042
	Average error	0.039	0.042	0.040	0.039	0.044	0.046	0.038	0.038	0.040

variables are correlation in model, that is, multicollinearity is existed in variables; to BPNN method, the maximal absolutee error of the results of the prediction and experiment is within 0.047mm, the average error is within 0.041mm, which can meet the needs of the engineering. However, since it can not be known exactly to conception and nature of approximation samples in BPNN application, this is very difficult even for reaching actual requirement even network error is 0. it is even hard to solve is replaced by a very small error. This is the so-called over fitting the phenomenon. It will effect directly generalization ability of network and will also make network lost their practical value at last. The generalization ability factors of effecting network are as follows,including sample properties and network itself influenced by many factors. Even though three methods can meet the needs of the engineering, through the comparison and analysis of the data of the experiment, No matter what you take into account, error or stability in the whole process monitoring, PLSR is more suitable for monitoring tool wear.

6. Conclusions

1. PLSR is suitable for analysis and modeling for monitoring tool wear in milling process. The results of the prediction and experiment on line show that tool wear mathematic model can obtain better tool wear value.
2. When monitoring tool wear is used to PLSR method, after selecting appropriate variables, existing multiply linearity correlates between dependent variable and first principal component of independent variables, PLSR model can obtain the high precision and reliability and its expression is simple and clear. It can give easy reasonable interpretation to factors effecting tool wear in cutting process and also can reveal the relationship between independent variables and dependent variable.
3. From computational requirement.Unfortunately, in most application the assumption is not reasonable. In order to make feature vectors more suit to be modeled, PLSR modeling may use the direct computation method and also iterative algorithm. Its calculation amount is small, at the same time calculation amount of prediction is very small. This is very important to research real-time on-line modeling and correct model algorithm.
4. Compared the regression model of the part sample variable with the model of the all sample variable, it can be found that the latter has better forecast effect than the former has, but the former can adopt to the polytrope on the coarse production easily.

Acknowledgements

The authors would like to thank the financial support from the Natural Science Foundation of Gansu Province.

References

Chen, J.C. (2004), "A multiple-regression model for monitoring tool wear with a dynamometer in milling operations", *Int. J. Technol. Studies*, **30**(4), 71-77.

Dimla, D.E. (2002), "The correlation of vibration signal features to cutting tool wear in a metal turning operation", *J. Mater. Process. Tech.,* **19**(10), 705-713.

Dragos, A. (2003), "Tool condition monitoring in broaching", *Int. J. Wear*, **254**(3), 370-382.

Erdogmus, D. (2004), "Minimax mutual information approach for independent component analysis", *Neural Comput.*, **16**(6), 1235-1252.

Ertunc, M.H. and Loparo, A.K. (2001), "Tool wear condition monitoring in drilling operations using hidden Markov models (HMPs)", *Int. J. Mach. Tool Manu.*, **41**, 1363-1384.

Kamarthi, S.V. and Kumara, S.R.T. (2000), "Flank wear estimation in turning through wavelet representation of acoustic emission signals", *J. Manuf. Sci. E. - T. ASME.*, **122**(1), 12-19.

Kopac, J. and Sali, S. (2001), "Tool wear monitoring during the turning process", *J. Mater. Process. Tech.*, **113**(1), 312-316.

Liangyue Cao. (1997), "Practical method for determining the minimum embedding dimension of a scalar time series", *Int. J. Phys. D*, **110**(1), 43-50.

Mannan, M.A. and Kassim, A. (2000), "Application of image and sound analysis techniques to monitor the condition of cutting tools", *Pattern. Recogn. Lett.*, **21**(11), 969-979.

Packard, N.H. (1980), "Geometry from a time series", *Int. J. Phys. Rev.*, **45**, 712-716.

Palanisamy, P. and Rajendran, I. (2007), "Prediction of tool wear using regression and ANN models in end-milling operation", *Int. J. Adv. Manuf. Tech.*, 1433-3015.

Shao, H. and Wang, H.L. (2004), "A cutting power model for tool wear monitoring in milling", *Int. J. Mach. Tool. Manu.*, **44**(14), 1503-1509.

Srinivasa, P. (2002), "Acoustic emission analysis for tool wear monitoring in face milling", *Int. J. Prod. Res.*, **40**(5), 1081-1093.

Tang, Q.Y. (2002), "Practical statistic analysis and DPS data processing system", *Science Press*, Beijing.

Wang, H.L. and Shao, H. (2003), "On-line tool breakage monitoring in turning", *J. Mater. Process. Tech.*, **139**(1), 237-242.

Wang, H.W. (2005), "The partial least-squares regression method and application", *National Industry Press*, Beijing.

Wang, S.B. and Cao, Y.D. (2005), "Deformation measurements using mutual information : Principle and applications", *Int. J. Optic. Laser. Eng.*, **43**(7), 895-903.

Xu, C.G. and Wang, X.Y. (2006), "New approach on monitoring cutting tool states by motor current", *J. Instrum. Technique Sensor*, **32**(4), 34-40.

Xu, C.W. (2003), "Design and research for tool wear monitoring", *Harbin Inst. Tech.*, **35**(1), 76-80.

Xu, C.W. and Chen, H.L. (2007), "A research of tool wear recognizing based on wavelet packet pretreated and neural network", *The International Journal of System Design and Dynamics, JPN Soc. Mech. Eng.*, **1**(4), 760-770.