www.s2is.org

International Journal on Smart Sensing and Intelligent Systems

ISSN 1178-5608

STUDY ON FEATURE SELECTION AND IDENTIFICATION METHOD OF TOOL WEAR STATES BASED ON SVM

Weilin Li, Pan Fu, Weiqing Cao

School of Mechanical Engineering

Southwest Jiaotong University, Er round Road 111#

Chengdu, Sichuan 610031, China

Emails: hiweilin@gmail.com; pfu@home.swjtu.edu.cn

Submitted: Jan. 24, 2013 Accepted: Mar. 12, 2013 Published: Apr. 10, 2013

Abstract- This paper presents an on-line tool wear condition monitoring system for milling. The proposed system was developed taking the cost and performance in practice into account, in addition to a high success rate. The cutting vibration signal is obtained during the cutting process, and then extracting features using time-domain statistical and wavelet packet decomposition algorithms. It would result in two major disadvantages if creating a tool wear states identification model based on all extracted features, i.e. high computational cost and inefficient complexity of the model, which leads to overfitting. It is crucial to extract a smaller feature set by an effective feature selection algorithm. In this paper, an approach based on one-versus-one multi-class Support Vector Machine Recursive Feature Elimination (SVM-RFE) is proposed to solve the feature selection problem in tool wear condition monitoring. Moreover, in order to analyze a performance degradation process on the machine tool, Least Squares Support Vector Machines (LS-SVM) is introduced. In order to estimate the

effectiveness of feature selection algorithm, the comparative analysis among Fisher Score (FS) Information Gain (IG) and SVM-RFE is exploited to real milling datasets. The identification result proves that: The selected feature set based on SVM-RFE is more effective to recognize tool wear state; LS-SVM wear identification method is superior to BP neural network, and it has higher identification accuracy; the proposed feature selection and identification method for tool wear states is efficient and feasible.

Index terms*:* Tool condition monitoring, feature selection, multi-class support vector machine recursive feature elimination (SVM-RFE), least squares support vector machines (LS-SVM).

I. INTRODUCTION

Condition monitoring system is very crucial to assure the reliability and safety of automatic, unmanned and adaptive machining. It is necessary to monitor the machining systems real-timely and accurately, especially to cutting tools [1]. Tool condition directly affects the quality, efficiency and safety of production. The tool change strategy is also change from the old practice, to the feasibility of instituting tool change procedures based on monitoring the amount of wear on the cutting tool-edges through the adaptive tool monitoring mechanisms. Many scholars, at home and abroad, have been doing many researches into tool wear monitoring technology. Roth John T [1], Teti R [2] and Abellan-Nebot [3] analyze the development states, trends and existing issues of tool condition monitoring system. The process of data-driven tool wear monitoring generally includes three steps: Monitoring signal acquisition; Feature extraction; Wear states identification.

The application of sensors encompasses many sectors of industry [4]. Tool wear monitoring has been performed using many different sensing techniques including cutting temperature, motor current, acoustic emission, vibration and force [5]. And among these sensing techniques, force sensor is most expensive [6], cutting vibration signal is free from the influence of chips and coolant, it is also very direct, real and low cost [7]. Meanwhile, the condition monitoring technology based on vibration signal is quite mature, and many achievements have been got in tool condition monitoring and fault diagnosis area. So the cutting vibration has been chose as sensing information in this paper. The features are extracted by time-domain statistical algorithms and wavelet packet algorithm. It is the aim of this work to integrate the vibration sensors to extract the largest possible effective information and got the tool wear by recognition model in cutting process.

Feature selection that is the process of adopting an appropriate subset of features out of the feature space is the key to the tool condition monitoring. An effective feature subset that is sensitive to tool wear is the basis of tool wear monitoring [2]. Usually, the original feature set is composed of the sensor information that reflects a specific dynamical behavior of the cutting tool, but some features are extracted as abstract mathematical representations without a particular physical meaning. Feature information in tool monitoring signals related to the discrimination of the tool wear states is therefore represented by a large dimensional space, which hinders the correct identification of the tool wear states. The known difficulties of large dimensionality spaces are: high processing times, complexity, and the well-known effect of the curse of dimensionality. It is necessary to extract a smaller feature set from high dimensional feature space with feature selection algorithm, which can improve the sensitivity of monitoring features. In recent years, many feature selection technologies for tool condition monitoring have been developed, which were reviewed in [2] and [8]. Peculiarly, under the framework of SVM, Jebara and Jaakkola [9] proposed the maximum entropy discrimination for feature selection. Weston etal [10] applied the gradient descent method to select good indicators under the given number of features. Support Vector Machine Recursive Feature Elimination (SVM-RFE) has become an attractive method for feature selection [11]. It gradually eliminates a variable whose removal changes the objective function used by the SVM the least in a sequential backward elimination manner. In order to extend binary SVM-RFE to multi-class SVM-RFE, several methods based on binary SVM have been proposed such as “one-versus-rest” (OVR), “one-versus-one” (OVO) [12]. This study uses an approach based on OVO-SVM-RFE for tool wear feature selection. With the help of SVM-RFE, the features which are irrelevant or less relevant to the tool wear will be deleted. The experimental result shows that the developed method can yield the more effective feature set in the tool wear identification.

Wear states identification modeling of performance degradation process is a critical procedure to estimate the tool wear. There is still shortage of satisfactory methods to model the performance degradation process when only few degradation samples are available. It is difficult of current performance degradation models to describe machining performance degradation of tool wear, which have small samples available. Neural networks (NNs) have been widely employed as the classifier to identify the tool state [8]. Compared with other learning algorithms, the SVM possesses a firm background and excellent features, such as minimizing the system complexity, yielding a significant gain in classification accuracy. SVM is a novel machine learning method based on statistical learning theory, which can be used for the classification and regression with small samples, nonlinearity, high dimension and local minima [13]. Least Squares Support Vector Machines (LS-SVM), derived from the standard SVM, has a better resistance impatient capacity and a faster operation speed [14]. Therefore, LS-SVM is introduced to analyze the process of machining performance degradation on the tool wear in this paper.

The rest of the paper is organized as follows. Section II gives a short description of two-class SVMs, OVO multi-class SVMs, and multi-class SVM-RFE feature selection method; Section III is about the experimental study on milling datasets for feature selection and wear identification. Section IV contains the conclusions.

II. RESEARCH METHOD

Tool wear monitoring system is composed of an accelerometer, data-acquisition devices and a micro-computer. The flank wear of cutting tool is the monitoring object. Multi-channel vibration signals are collected and converted to digital signals to feed into the computer which will accomplish data processing. Figure 1 is the schematic diagram of the milling tool wear states monitoring system.

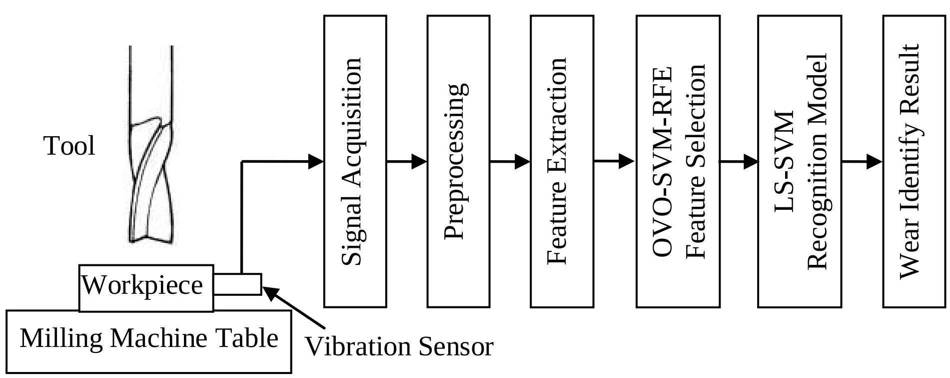

Figure 1. Schematic diagram of tool wear monitoring system

a. Two-Class and multi-class SVMs algorithms

a.i Two-Class SVMs

SVM were developed by Vapnik and his colleagues within the context of statistical learning theory and structural risk minimization [15]. The basic concept of SVM is to transform the input vectors to a higher dimensional space by a nonlinear transform, and then an optical hyperplane which separates the data can be found. This hyperplane should have the best generalization capability.

For training data set $(x_1, y_1),...,(x_l, y_l), y_i \in \{-1,1\}$, to find the optical hyperplane *H*, a nonlinear transform, $Z = \Phi(x)$, is applied to *x*, to make *x* become linearly dividable. A weight w and offset b satisfying the following criteria will be found [16]:

$$\begin{cases} w^T z_i + b \geq 1, & y_i = 1 \\ w^T z_i + b \leq -1, & y_i = -1 \end{cases} \tag{1}$$

Assume that the equation of the optical hyperplane *H* is $w_0{}^T z + b_0 = 0$, and then the distance of the data point in any of the two classes to the hyperplane is:

$$\rho(w,b) = \min_{x|y=1} \frac{z^T w}{\| w \|} - \max_{x|y=-1} \frac{z^T w}{\| w \|} \tag{2}$$

Then the search of the optimal plane *H* turns to a problem of a second order planning problem.

$$\min_{w,b} \Phi(w) = \frac{1}{2}(w^T w) \qquad (3) \qquad \text{Subject to } y_i(w^T z_i + b) \geq 1, \quad i = 1, 2, ..., l \qquad (4)$$

By using Lagrange method, the decision function of

$$w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i \qquad (5) \qquad f = \text{sgn}[\sum_{i=0}^{l} \lambda_i y_i (z^T z_i) + b] \qquad (6)$$

From the functional theory, a non-negative symmetrical function $K(x_i, x)$ uniquely define a Hilbert space *H*. This stands for an internal product of a characteristic space:

$$z_i{}^T z = \Phi(x_i)^T \Phi(x) = K(x_i, x) \tag{7}$$

Then the decision function can be written as:

$$f = \text{sgn}[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b] \tag{8}$$

a.ii Multi-class SVMs

SVM is designed to solve a binary classification problem. For tool-wear feature selection problem, which is a multi-class problem, classification is accomplished through combinations of binary classification problems. There are two ways to do that: one vs. one (OVO) or one vs. all (OVA), which are the basis of the multi-class feature selection method to be presented.

OVO-SVM [12]: Given a classification problem with M classes, OVO-SVM constructs $M(M-1)/2$ binary SVM classifiers, each for every distinct pair of classes. Each of binary classifiers takes samples from one class as positive and samples from another class as negative. Max-Wins voting (MWV) is one of the most commonly used combination strategies for OVO-SVM. MWV assigns an instance to a class which has the largest votes from $M(M-1)/2$ OVO binary classifiers.

OVA-SVM [12]: OVA-SVM constructs M binary classifiers and each binary classifier classifies one class (positive) versus all other classes (negative). Winer-Takes-All (WTA) is the most common combination strategy for OVA-SVM. WTA strategy assigns a sample to the class whose decision function value is largest among all the M OVA binary classifiers.

b. Multi-class SVM-RFE for feature selection

SVM-RFE [17] as a feature selection method for two-class classification was initially developed for gene selection. Features are eliminated one by one in a backward selection procedure that is referred to as Recursive Feature Elimination (RFE). Once a feature is removed, a new linear two-class SVM is trained with all the features left and all the remaining features are ranked again by using weight vector of the new linear SVM. SVM-RFE repeats this procedure until only one feature is left.

In this study, the binary SVM-RFE was extended to multi-class SVM-RFE. OVO-SVM were proposed for multi-class classification with feature ranking scores computed from weight vectors of multiple binary SVM classifiers. The nonlinear RBF kernel function was preferred in this study to deal with the nonlinear relationship between product form features.

For a multi-class problem with M classes, suppose T nonlinear binary SVM classifiers are obtained from a OVO-SVM or OVA-SVM multi-class classifier; $T=M(M-1)/2$ for OVO-SVM and $T=M$ for OVA-SVM. Let w_j be the weight vector of the *j-th* linear SVM and w_{ji} be the corresponding weight value associated with the *i-th* feature; let $v_{ij} = (w_{ji})^2$. We can compute feature ranking scores with the following criterion:

$$c_i = \frac{\frac{1}{T}\sum_{j=1}^{T} v_{ij}}{\sqrt{\frac{\sum_{j=1}^{T}\left(v_{ij}-\overline{v}_i\right)^2}{T-1}}} \qquad (9)$$

Where $\overline{v}_i$ are mean of variable v_i.

With feature ranking criterion, features are selected using the following backward elimination procedure similar to that of SVM-RFE:

(1) Start with an empty ranked features list *R= []* and the selected feature list *F= [1, ··; d]*;

(2) Repeat until all features are ranked:

1) Train *n(n-1)/2* or *n-1* SVMs with all the training samples, with all features in *F*;

2) Compute and normalize the weight vectors;

3) Compute and sum the ranking scores of SVMs for features in *F* using Eq. (9);

4) Find the feature with the smallest ranking criterion: $e = argmin_f\, c_f$;

5) Add the feature *e* into the ranked feature list *R: R= [e, R]*;

6) Remove the feature *e* from the selected feature list *F: F=F-[e]*;

(3) Output: Ranked feature list *R*.

Refer to the above proposed multi-class SVM-RFE with binary classifiers from OVO-SVM and OVA-SVM as OVO-SVM-RFE and OVA-SVM-RFE. All features are ranked from the RFE procedure. Earlier one feature is eliminated, lower is it ranked. Thus, nested feature subsets are obtained similarly as in filtering feature selection methods which however usually rank all features at a single step.

c. LS-SVM for wear identification

LS-SVM are least squares versions of support vector machines (SVM), which are a set of related supervised learning methods that analyze data and recognize patterns, and which are used for classification and regression analysis.. Comparing with simple SVM, Only linear equation is needed to solve the results, which avoids the local minima in SVM. A short summary of the LS-SVM is given here; more details are given in [18].

The LS-SVM model is defined in its primal weight space by [19]:

$$\hat{y} = w^T \phi(\mathbf{x}) + b \qquad (10)$$

Where $\phi(\mathbf{x})$ is a function which maps the input space into a higher dimensional feature space, x is the N-dimensional vector, w and b the parameters of the model. In LS-SVM for function estimation, the following optimization problem is formulated:

$$\min_{w,b,e} J_{LS}(w,b,e) = \tfrac{1}{2} w^T w + \gamma \tfrac{1}{2} \sum_{k=1}^{N} e_k^2 \tag{11}$$

Subjects to the equality constraints

$$y_k[w^T \phi(x_k) + b] = 1 - e_k, \quad k = 1,...,N \tag{12}$$

The Lagrangian is defined as

$$L(w,b,e;\alpha) = J_{LS} - \sum_{k=1}^{N} \alpha_k \left\{ y_k[w^T \phi(x_k) + b] - 1 + e_k \right\} \tag{13}$$

With Lagrange multipliers $\alpha_k \in R$ (called support values).

The conditions for optimality are given by

$$\begin{cases} \frac{\partial L}{\partial w} = 0 & \rightarrow & w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k) \\ \frac{\partial L}{\partial b} = 0 & \rightarrow & \sum_{k=1}^{N} \alpha_k y_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 & \rightarrow & \alpha_k = \gamma\, e_k \\ \frac{\partial L}{\partial \alpha_k} = 0 & \rightarrow & y_k[w^T \phi(x_k) + b] - 1 + e_k = 0 \end{cases} \tag{14}$$

For $k=1, 2, \cdots; N$. After elimination of w and e one obtains the solution

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \tag{15}$$

With $Z = [\phi(x_1)^T y_1;...;\phi(x_N)^T y_N]$, $Y = [y_1;...;y_N]$, $1_v = [1;...;1]$, $e = [e_1;...;e_N]$ and $\alpha = [\alpha_1;...;\alpha_N]$.

Where α and b are obtained by the solution.

Mercer's condition is applied within the matrix ZZ^T, and then we have:

$$\mathrm{ZZ}^{\mathrm{T}} = y_i y_j K\left(x_i, x_j\right) \tag{16}$$

Where $K(x_i,x_j)$ is the kernel function.

The Radial Basis Function (RBF) kernel can reduce the computational complexity of the training process and thus improve the LS-SVM performance. Therefore, RBF kernel is selected as kernel function in tool wear identification model.

According to the LS-SVM regression algorithm, the regularization parameter γ and RBF kernel parameter σ play an important role to the assessment results from the posterior reliability assessment model. In the paper, these parameters were determined as follows [20] [21]: First, Coupled Simulated Annealing (CSA) determines suitable starting points for every method. Second, these parameters are then given to a second simplex optimization procedure to perform a fine-tuning step.

III. EXPERIMENTAL RESULTS

a. Experimental design

The experimental data sets were provided by PHM Society [22]. Experiments are carried out on a high speed CNC machine (Röders Tech RFM760) with spindle speed up to 42,000 rpm. The workpiece material was stainless steel (HRC52). The surfaces of workpieces were prepared through face milling to get rid of the original skin layer. The surface was then machined to have a slope with 60° to accommodate the 3-flute ball nose cutter (*Figure 2*). Three Kistler piezo accelerometers were mounted on the workpiece to measure the machine tool vibrations of cutting process in X, Y, Z direction respectively. The amplified voltage signals were captured by a NI DAQ PCI 1200 board with 50 kHz/channel frequency.

The experimental conditions are as follows: The spindle speed is 10400 r/min; the feed rate is 1555 r/min; the Y-axis and Z-axis cutting depth is 0.125 mm and 0.2mm; the length of each face milling is 108 mm; the tool flank wear that will be measured by microscope after each face milling is used as the tool wear value. Tool wear was divided into 60 levels from 0.091mm~0.150mm with the resolution ratio of 0.001mm. There are 8 samples in each wear levels, with 4 samples were used for model training and 4 samples were used for model test.

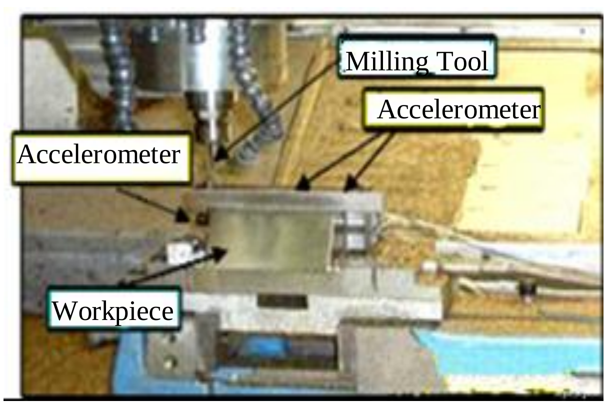

Figure 2. CNC milling machine testbed

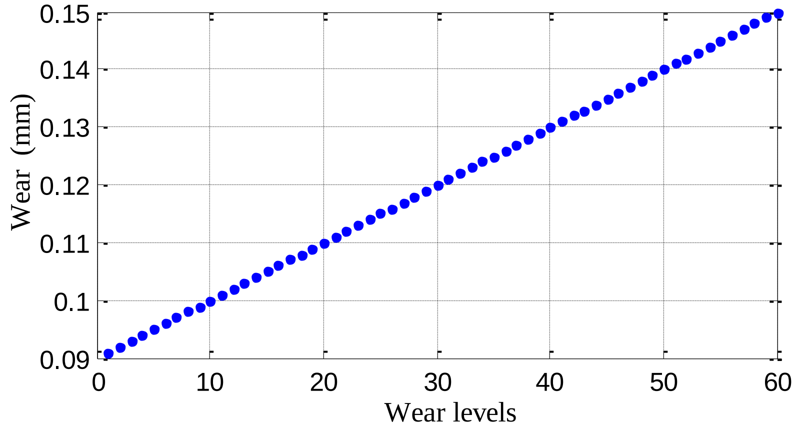

Figure 3. Experiment tool wear levels

b. Feature extraction

Although vibration signal might be sufficiently available on the entire machining process, the main drawback of this signal is its both sensitive to variation and noise from cutting operation. Figure 4 is time-domain vibration waveform in X axis. Therefore, the feature extraction is a key issue of tool condition monitoring, which requires a powerful signal processing to maximize the information utilization of the sensor signals. In order to eliminate the redundancy and highlight the condition information, the relative statistical features of time and frequency domain are extracted.

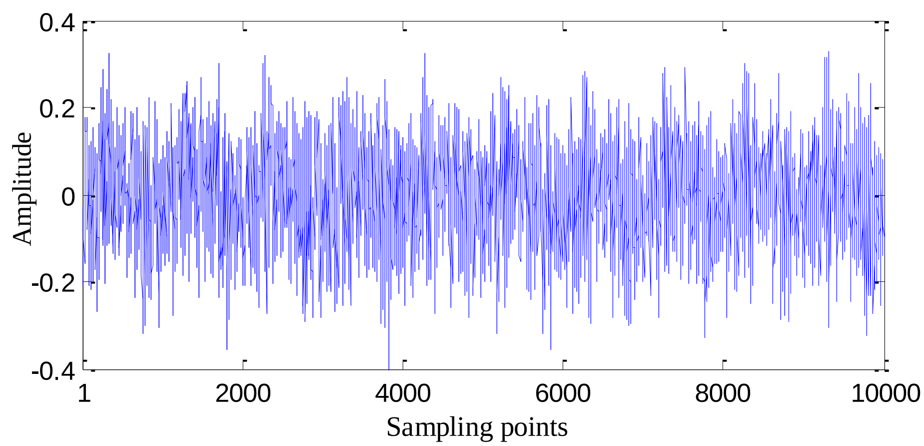

Figure 4. Time-domain vibration waveform in *X* axis

In this paper, there are 7 time-domain statistical features of tool condition monitoring in each acquisition channels including variance, peak, root mean square, square root value, kurtosis, crest, and shape factor.

Since there is rich information of tool states in frequency-domain, frequency domain feature such as energies of the spectral on divided frequency bands also have been extracted. With wavelet packet transform of tool wear signals, energies of divided frequency bands could be calculated. After i-level wavelet packet decompositions, the energy of the j-th node in i-th level can be computed as:

$$E_{ji} = \sum_k \left(r_i \left(k \right) \right)^2 \tag{18}$$

where *ri=(ri(1), ri(2), ··; ri(n))* is the reconstructed signal obtained from the wavelet packet coefficients corresponding to different frequency rang.

In this paper, a 7-level Daubechies wavelet packet transform is used to calculate the frequency bands energy of tool monitoring signal, and 128 frequency bands were got in each acquisition channels. Because the energy of high-frequency noise is small and has no contribution to the tool state identification, only 1-64 band energies are used as frequency-domain features. There are three vibration acquisition channels (X, Y, Z direction) and 71 features in each direction (7 time-domain and 64 frequency-domain), so the 71×3=213 normalized features constructed the original feature space of tool wear monitoring, which are described in Table1

Table1: Index of original features

Index	**Definition**
1-64	64 frequency bands energy of X direction
64-128	64 frequency bands energy of Y direction
129-192	64 frequency bands energy of Z direction
193-199	7 time-domain statistical features of X direction
200-206	7 time-domain statistical features of Y direction
217-213	7 time-domain statistical features of Z direction

These features cannot be intuitively compared and decided which one is more suitable and reliable to be used as monitoring indices or as an input feature vector to decision algorithm, since most of them fluctuate greatly in the whole machining process(*as shown in Figure 5*). In the following section, the selection method of these features will be discussed.

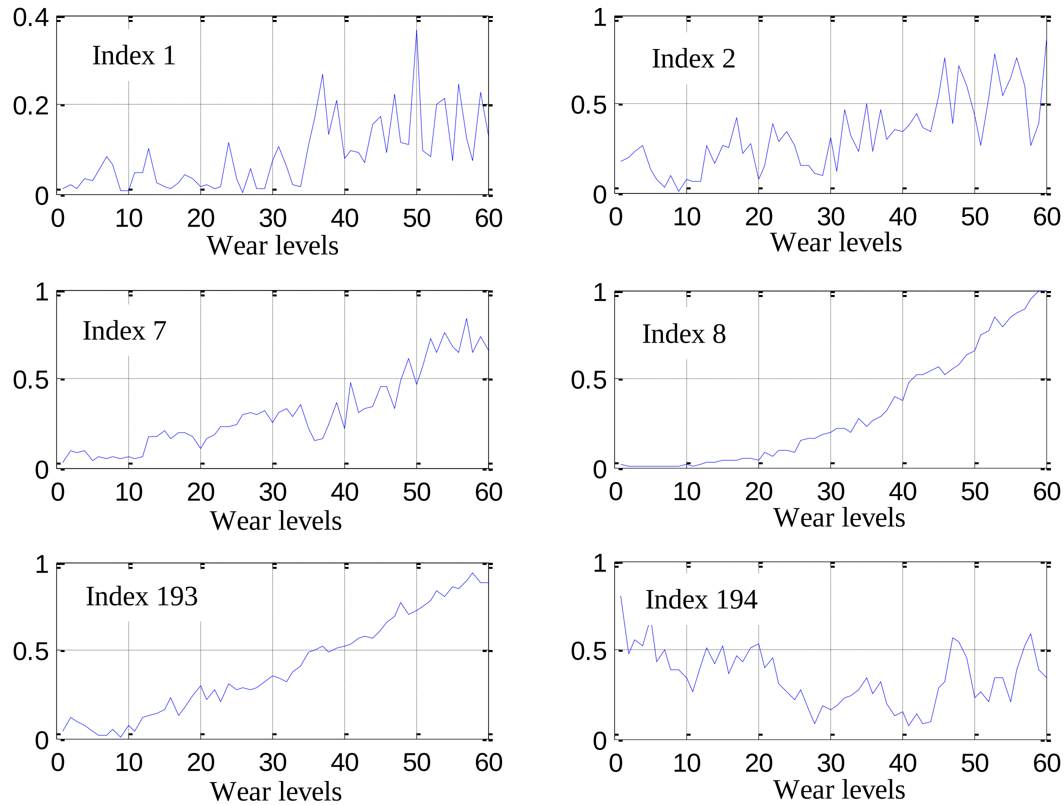

Figure 5. Parts of features of *X* direction in different wear level

c. Feature selection and wear identification

The proposed approach aims to construct a smaller feature set form original features by selection model based on OVO-SVM-RFE. First, an OVO multiclass SVM model using a RBF kernel was constructed; each tool-wear sample was assigned a class label and formulates a multiclass classification problem divided into a series of OVO SVM sub-problems; each test sample was sequentially presented to each of the $n(n-1)/2$ OVO classifiers (there are 60 wear levels, n=60). Then, a multiclass SVM-RFE process was conducted to select the critical features. The relative importance of the features can be analyzed during each iterative step.

To quantify the performance of our algorithm in a comparative manner, other two feature selection methods were used:

Fisher Score (FS): The fisher score is a method for determining the most relevant features for classification. It uses discriminative methods, and generative statistical models to accomplish this. Fisher Score is an effective supervised feature selection algorithm, which has been widely applied in many real applications.

Information Gain (IG): Information Gain is a measure of dependence between the feature and the class label. It is one of the most popular feature selection techniques as it is easy to compute and simple to interpret.

After selecting the features, LS-SVM is then performed to describe the relationship between the monitoring information and the tool wear states by only using the selected features. In this case, the input vector for LS-SVM in experiment is a subset of the original features selected by OVO-SVM-RFE; the output is the tool flank wear (*VB*). RBF kernel is selected as kernel function in posterior reliability assessment model. The regularization parameter γ and RBF kernel parameter σ were obtained using Coupled Simulated Annealing (CSA) and simplex technique. Once the LS-SVM has learned the correlation between the input information and tool flank wear, it can be used to predict the tool condition. Identification error was selected as criteria of feature usability assessment. There are 60 wear levels with 4 test samples in each level, so there are $60\times4=240$ test samples in all.

Figure 6 gives a comparison of identification performances in different feature selection dimension by OVO-SVM-RFE, FS and IG on milling datasets. When the selection features dimension was only 8 (*Point 1 in figure 6*), the identification accuracy of SVM-RFE and FS are close to the result of using all 213 original features, which indicates the efficient of feature selection algorithm. When the selection features dimension was 20 (*Point 2 in figure 6)*, SVM-RFE can get the highest accuracy in all dimension compare with FS and IG. The identification error increases as the dimension increases when the dimension is more than 20, so 20 was the final dimension of tool wear features selection.

Meanwhile, OVO-SVM-RFE algorithm performs especially well when the number of selected features is small from figure 6. Comparing to FS and IG, SVM-RFE can achieve much more compact representation without sacrifice of discriminating power.

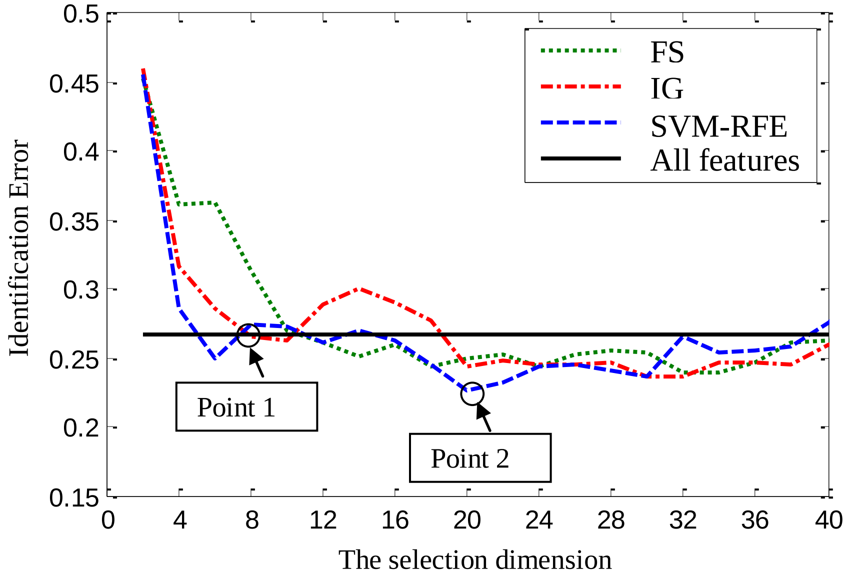

Figure 6. Performance of selected feature subset

When the dimension of tool wear features selection is 20, the rank index of feature subset that is ranked from smallest to largest is list in Table 2. In SVM-RFE, 8, 20, 28, 52 belong to the bands energy of *X* direction; 68, 72, 82, 84,119 belong to ands energy of *Y* direction; 131, 135, 156, 167, 180 belong to ands energy of *Z* direction; 193 belong to time-domain statistical features of *X* direction; 200, 201, 202, 206 belong to time-domain statistical features of *Y* direction; 207 belong to time-domain statistical features of *Z* direction.

Table 2: List of selected features

Rank index of SVM-RFE																			
8	20	28	52	68	72	82	84	119	131	135	156	167	180	193	200	201	202	206	207
Rank index of FS																			
7	8	18	20	39	52	68	72	82	84	104	114	156	167	180	193	195	200	202	207
Rank index of IG																			
7	8	18	20	37	40	45	52	68	72	84	104	114	116	156	180	193	200	202	207

Finally, a comparison between LS-SVM and artificial neural network (ANN), in particular, the BP neural network, was made to estimate the tool condition with the monitoring information. For this comparison, the selected features based on SVM-RFE were used.

Figure 7 gives the results of the LS-SVM estimation for the optimal values of $\gamma=2011$ and $\sigma^2=129$. Figure 8 gives the results of the network whose architecture was 20-41-1 network based on minimum RMS error. Comparing the results of these two methods, the error of the ANN is greater than the error obtained with LS-SVM. In other words, it seems that LS-SVM presents a greater capability of generalization than the ANN.

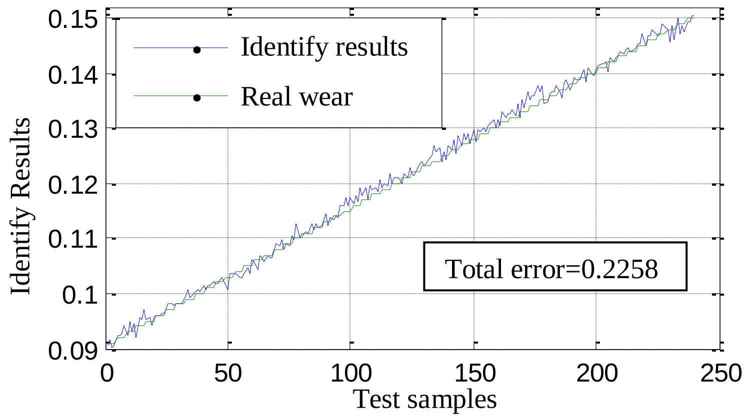

Figure 7. Performance of feature subset based on LS-SVM

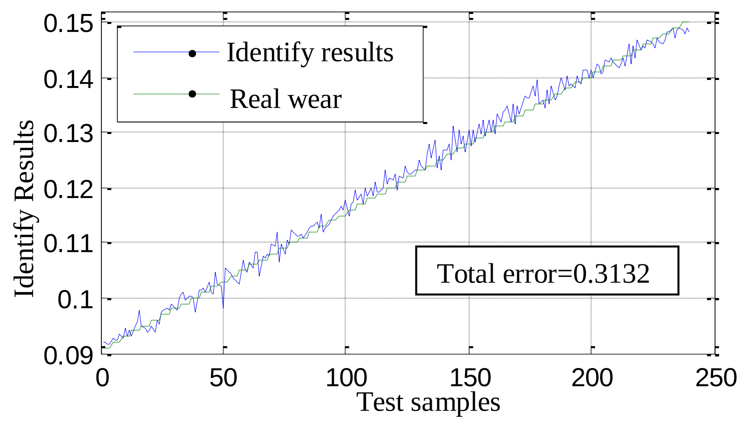

Figure 8. Performance of feature subset based on BP neural network

VI. CONCLUSIONS

Feature selection and identification method are the key to the tool condition monitoring. Multi-class SVM-RFE and LS-SVM was proposed to deal with these two issues. First, the features are extracted from vibration monitoring signal by time-domain statistical algorithm and wavelet packet decomposition algorithm. Second, a more sensitive feature subset was selected from the original features by OVO-SVM-RFE. Last, a tool wear identification model based on LS-SVM was built, and tool wear can be effectively identified. The validity of this method has been proved by the milling tool monitoring experiment.

OVO-SVM-RFE can well handle tool wear multi-class feature selection problem. The feature selection process can not only reduce the cost of recognition by reducing the number of features that need to be collected, but also improve the classification accuracy of tool condition monitoring system. In comparison with Fisher Score and Information Gain, the experimental results validate that the OVO-SVM-RFE method achieves significantly higher performance, which can performs especially well when the dimension of selected features is 20.

LS-SVM is a superior method for tool wear identification and can get satisfactory results for tool wear identification. Experiment results show the validity and practicability of the method. In particular, LS-SVM presents a good estimation error and a greater capability of generalization in the comparison with BP neural network.

ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities (SWJTU12CX039).

REFERENCES

[1] Roth John T, Djurdjanovic Dragan, Yang Xiaoping, “Quality and Inspection of Machining Operations: Tool Condition Monitoring”, Journal of Manufacturing Science and Engineering, Vol.132, No.4, pp. 0410151-04101516, 2010.

[2] Teti R, Jemielniak K, O'Donnell G, Dornfeld D, “Advanced Monitoring of Machining Operations”, CIRP Annals—Manufacturing Technology, Vol. 59, No. 2, pp. 717-739, 2010.

[3] Abellan-Nebot, Jose Vicente, Romero Subirón Fernando, “A Review of Machining Monitoring Systems Based on Artificial Intelligence Process Models”, The International Journal of Advanced Manufacturing Technology, Vol. 47, No. 1-4, pp. 237-257, 2010.

[4] T.Jayakumar, C.Babu Rao, John Philip, C.K.Mukhopadhyay, J.Jayapandian, C.Pandian, “Sensors for Monitoring Components, Systems and Processes”, International Journal on Smart Sensing and Intelligent Systems, Vol. 3, No. 1, pp. 61-74, March 2010.

[5] E. Dimla Snr, “Sensor Signals for Tool-Wear Monitoring in Metal Cutting Operations—A Review of Methods”, International Journal of Machine Tools & Manufacture, Vol. 40, pp. 1073–1098, 2000.

[6] Boukhenous, S., “A Low Cost Three-Directional Force Sensor”, International Journal on Smart Sensing and Intelligent Systems, Vol. 4, No. 1, pp. 21-34, March 2011.

[7] Li Weilin, Fu Pan, Cao Weiqing, “Tool Wear States Recognition Based on Frequency-Band Energy Analysis and Fuzzy Clustering”, In Proceedings of 3rd International Workshop on Advanced Computational Intelligence (IWACI 2010), pp. 162-167, 2010.

[8] Bernhard Sick, “On-Line and Indirect Tool Wear Monitoring in Turning with Artificial Neural Networks: A Review of More Than A Decade of Research”, Mechanical Systems and Signal Processing, Vol.16, No. 4, pp. 487–546, July 2002.

[9] Tony Jebara and Tommi Jaakkola, “Feature Selection and Dualities in Maximum Entropy Discrimination”, Proceedings of the Sixteenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000), pp. 291-300, 2000.

[10] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, “Feature Selection for SVMs”, in Proc. NIPS pp.668-674, 2000,.

[11] I. Guyon, J. Weston, S. Barnhill, and N. Vapnik, “Gene selection for Cancer Classification Using Support Vector Machines”, Machine Learning, Vol. 46, no. 1-3, pp. 389–422, 2002.

[12] K-B. Duan, J.C. Rajapakse, and M.F. Nguyen, "One-Versus-One and One-Versus-All Multi-Class SVM-RFE for Gene Selection in Cancer Classification", Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Lecture Notes in Computer Science, Vol. 4447, pp. 47-56, 2007.

[13] L.J. Cao, F.E.H. Tay, "Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting", IEEE Transactions on Neural Networks, Vol. 14, No. 6, pp.1506–1518, 2003.

[14] I. Goethals, K. Pelckmans, J.A.K. Suykens, Bart De Moor, "Subspace identification Of Hammerstein Systems Using Least Squares Support Vector Machines", IEEE Transactions on Automatic Control, Vol. 50, No.10, pp.1509–1519, 2005.

[15] Corinna Cortes, Vladimir Vapnik, "Support-vector networks", Machine Learning, Vol 20, No. 3, pp. 273-297, September 1995.

[16] Nello Cristianini, John Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods", Cambridge University Press, 2000.

[17] Guyon, I., Weston, J., Barnhill, S., Vapnik, V. "Gene Selection for Cancer Classification Using Support Vector Machines", Machine Learning, Vol. 46, No.1-3, pp. 389–422, 2002.

[18] J.A.K. Suykens, J. Vandewalle, "Least Squares Support Vector Machine Classifiers", Neural Processing Letters, Vol. 9, Vol. 3, pp. 293-300, 1999.

[19] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, "Least Squares Support Vector Machines", World Scientific Pub. Co., Singapore, 2002.

[20] Xavier de Souza, S. Suykens, J. A. K. "Coupled Simulated Annealing", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 40, No. 2, pp. 320–335, 2010.

[21] Nelder J. A., Mead R., "A Simplex Method for Function Minimization", Computer Journal, Vol. 7, pp. 308-313, 1965.

[22] http://www.phmsociety.org/