Int J Adv Manuf Technol (2018) 95:775–784
https://doi.org/10.1007/s00170-017-1242-9

ORIGINAL ARTICLE

Online chatter detection of the end milling based on wavelet packet transform and support vector machine recursive feature elimination

G. S. Chen [1] · Q. Z. Zheng [1]

Received: 15 March 2017 / Accepted: 24 October 2017 / Published online: 3 November 2017

Abstract Chatter is a common state in the end milling, which has important influence on machining quality. Early chatter detection is a prerequisite for taking effective measures to avoid chatter. However, there are still many difficulties in the feature extraction of chatter detection. In this article, a novel online chatter detection method in end milling process is proposed based on wavelet packet transform (WPT) and support vector machine recursive feature elimination (SVM-RFE). The measured vibration signal in the machining process was preprocessed by WPT. The original feature set of chatter composed of ten time-domain and four frequency-domain feature parameters was obtained via calculating the reconstructed signal. Then feature weights are computed by SVM-RFE, and the obtained feature ranking list was to indicate their different importance in chatter. The optimal feature subset was selected according to the prediction accuracy. The proposed method is described and applied to incipient chatter over conventional methods in identifying the transition from a stable to unstable state. Some milling tests were conducted and the experiment results was shown that the impulse factor and onestep autocorrelation function were the sensitive chatter features.

Keywords Chatter detection · WPT · SVM-RFE · Feature parameters

✉ Q. Z. Zheng
zqz513@163.com

[1] School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Chatter is a common adverse phenomenon in the process of high-speed milling. It is a self-excited vibration between cutting force and vibration of tool-workpiece system. The occurrence of chatter has negative effects on poor surface quality, unacceptable inaccuracy and reduced material removal rate, seriously destroying the cutting tools and shortening the lifetime of machine tool. Some researchers reviewed about the chatter problems which is chatter prediction, chatter detection, and chatter control strategies [1–3]. However, considering the complexity of the chatter mechanism and cutting conditions, the chatter prediction is difficult to carry out in industrial production; therefore, chatter detection becomes crucial. The most reliable approach is to establish an automatic online detection system which is a prerequisite for maintaining efficient machining process and the realization of control strategies for chatter suppression. Hence, in order to improve the accuracy and workpiece quality, online chatter detection is the key factor to suppress the chatter, which has become the key research contents.

In the recent years, for the purpose of monitoring the cutting status, many chatter monitoring techniques have been developed through monitoring a certain signal sensor such as accelerometer [4], AE sensor [5], current [6], microphone [7], dynamometer [8, 9], and multisensors [10] to obtain the process information. No matter which signal is selected, the method of signal analysis is extremely important. Unfortunately, due to the measured vibration signal during the manufacturing process containing the background noise, which may not be directly used for chatter identification, it needs to be analyzed and extracted, the effective and sensitive features [11]. Therefore, appropriate method of feature extraction and feature selection should be found for chatter detection via the signal processing, which is crucial. Advanced monitoring and detection chatter methods are developed mostly depending on frequency and time-frequency analysis. Fast

Fourier transform (FFT) can convert signals from time domain to frequency domain, which can reflect the overall statistical characteristics of vibration signals. It is often used to detect cutting chatter [12]. It only has a good resolution in the frequency domain, but it masks the time-domain information. However, the measured vibration signal in machining processes is normally non-stationary and non-linear; the application of FFT is limited for online detection of chatter onset. Wavelet analysis expands its range of application because of its strong ability of local analysis. Somkiat et al. [13, 14] established online chatter detection of the ball milling based on wavelet analysis, which achieved good performance. However, when the chatter frequency occurs at a high frequency, the application of the high frequency signal is greatly limited because of the low resolution of the decomposed high frequency coefficient. To solve this issue, wavelet packet transform (WPT) is a very suitable choice, which can simultaneously decompose the low frequency and high frequency coefficients, and adaptively determine the resolution of different frequency segments in the signal. Yuxin Sun et al. [15] used the weighted wavelet packet entropy method to achieve online chatter detection of the turning process. Yao et al. [16] adopted the standard deviation of wavelet analysis and energy ratio of wavelet packet transform as the feature vector, which realized the turning machining condition accurately by the SVM. Chen Bing [17] used multi-scale permutation entropy and wavelet packet energy as the milling chatter premonition features, and the selected chatter features is depended on experimental analysis. However, the above selected chatter features, constructed on the basis of individual experience, are not always efficient to detect a defect at its early stage of chatter. Lamraoui [18] used the wiener filter to extract various statistical features for detect chatter in CNC milling process. The adepts of data fusion take advantage of a mass of features for effective condition detection. However, those approached would impact on the training time and classification accuracy owning to the presence of irrelevant or redundant features [19].

In order to improve the accuracy of chatter recognition and reduce the calculation time, it is more crucial to select the most sensitive features from the extracted features to identify chatter. The main purpose of chatter detection is to analyze the relevant external information in order to judge the condition of the inaccessible internal components so as to decide if the machine needs to be dismantled or not [20]. There are many feature selection methods. Li Sheng [21] combined the genetic algorithm and the partial least-squares method to select the characteristic of hydraulic system fault diagnosis, which shortened the calculation time and improved the classification accuracy. Li Weihua [22] adopted principal component analysis to select the fault characteristics in the gearbox early fault diagnosis. Lamraoui [19] applied a multiband resonance filtering to preprocess the vibratory signal before generated features. Extracted features were ranked based on their entropy in which only best features are selected. Then two neural network approaches, radial basis function and multi-layer perceptrons, classify the selected features into stable or unstable classes. Although these feature selection approaches have achieved superb performance, they have unstable and overfitting phenomena. Therefore, it is imperative to find a reliable and stable feature selection method. Support vector machine recursive features eliminating (SVM-RFE) is an embedded method of feature selection [23]. SVM as a small sample learning machine, it has many unique advantages in non-linear and high-dimensional pattern recognition, owning maximized generalization ability, and minimized classification errors [24]. SVM-RFE uses the weights of support vector machine as the evaluation criterion of feature selection, and only one feature is eliminated for each iteration. This method has a very effective effect in gene selection [25], signal processing [26], and other fields. Therefore, the method of combining wavelet packet transform and SVM-RFE used to detect chatter onset in high-speed milling may be feasible. Shicai Qian [27] utilized LSSVM-RFE to select the effective wavelet packet node energy feature online detection of chatter vibration. The experiment results were shown that the novel method can not only improve the accuracy of chatter recognition, but also decreased computing time.

Based on the above analysis, this paper proposed a novel approach to establish the online chatter detection of milling, which is combined WPT and SVM-RFE. Firstly, the vibration signal is preprocessed with WPT. The wavelet packets around twice the natural frequencies of system are selected and reconstructed when chatter occurs. Then, the ten time-domain and four frequency-domain features of reconstructed signal are calculated as the original feature set of chatter identification. The extracted features are ranked based on the algorithm of SVM-RFE, in which the optimal features are selected. Finally, the optimal feature subset was selected according to the prediction accuracy, which was fed into SVM for training and testing. The experiment results was shown that the selected best features were the sensitivity of features, which was a promising approach to detect chatter in end milling at an early stage.

1 Experimental setup

As shown in Fig. 1, the experimental machine is VMC1165B three-axis milling machine. The accelerometer was mounted on the spindle housing to collect the vibration signal during the milling process. The vibration signals were sampled by a data acquisition card, and then transmitted to the PC that was applied to save data and analyze signals.

During the cutting tests, a carbide end mill cutter with two flutes was applied to cut aluminum 6061. Besides, the cutting tool's diameter is 8 mm and the tool overhang is 44 mm. The

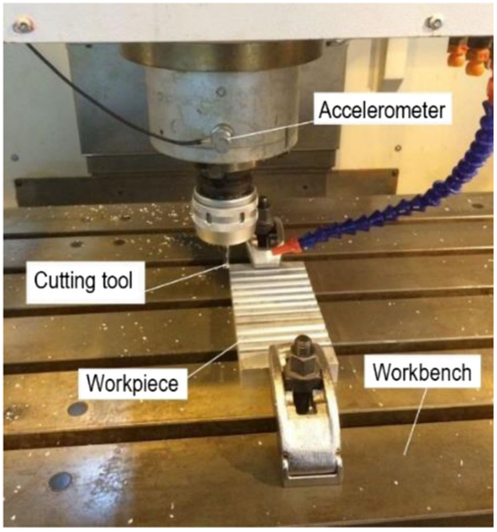

Fig. 1 Experimental setup

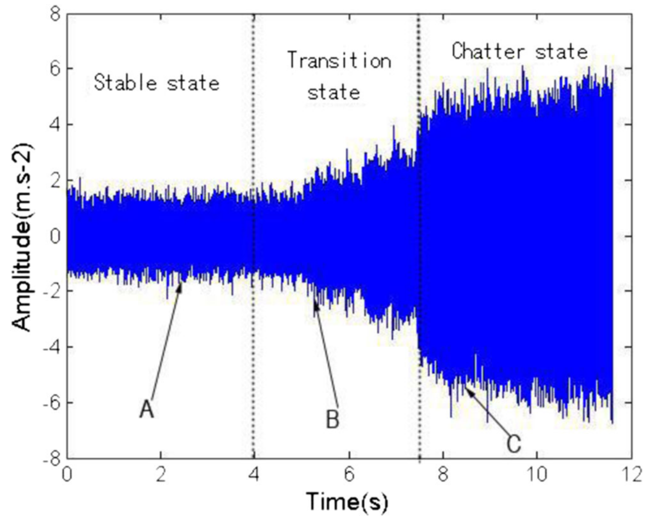

Fig. 2 Vibration signal of time domain

signal sampling frequency is set to 12,000 Hz. All tests are conducted without coolant.

Since, it was well known that chatter can arise at certain combinations of axial depth of cut and spindle speed during a milling process [19, 28]. Therefore, for the validation of the proposed method in this paper, the feed rate was the same (0.02 mm/per tooth). When the spindle speed is fixed, the cutting depth starts from 0.2 mm and increases 0.2 mm each time until chatter occurs. The other cutting parameters in this experiment are shown in Table 1.

Experimental conditions for the feed rate 0.02/z, spindle speed 7000 r/min and the depth of cutting 0.4 mm, the measured vibration signal is shown in Fig. 2. From Fig. 2, the occurrence of chatter is an energy accumulation process during the milling. There are three states in the milling process, which are stable state, transition state, and chatter state. It is seen that when chatter occurs, the amplitude of the vibration signal will increase significantly. At this moment, the surface quality and geometry of workpiece may have been damaged seriously. When the stable state is transformed into the chatter state during the milling process, there is a transition state where the amplitude of vibration signal does not increase significantly, but it has been pregnant with chatter onset [16].

Normally, the preliminary experiments have been conducted to examine the measured vibration signal and their FFT to check the chatter frequency. However, the chatter frequency is different since the chatter is influenced by the cutting parameters and the modal parameters of the machining system. Hence, a hammer test is applied on the tool tip to obtain the modal parameters of the spindle (including tool and tool holder) before the milling tests. A third orders natural frequencies are 1494, 2041, and 4160 Hz.

Figure 3 shows the measured vibration signal and their FFT, where a, b, and c are corresponding to A, B, and C of Fig. 2, respectively. From the figure, when chatter occurs during in the milling, the distribution of the frequency components has been changed. In Fig. 3b, the chatter frequency has obviously emerged, which is about 2719 Hz. The chatter frequency is close to the twice first natural frequency of the system [29] because the helix angle of the mill can have an important role on instability due to repetitive impact driven chatter [30]. And the amplitude also is slight increased. Therefore, the recognition of the transition stage becomes

Table 1 Cutting conditions

Workpiece	AL6061
Tool	Carbide End Mills $\Phi 8mm$
Spindle speed(rpm)	3000, 4000, 5000, 6000, and 7000
Feed per tooth(mm/z)	0.02

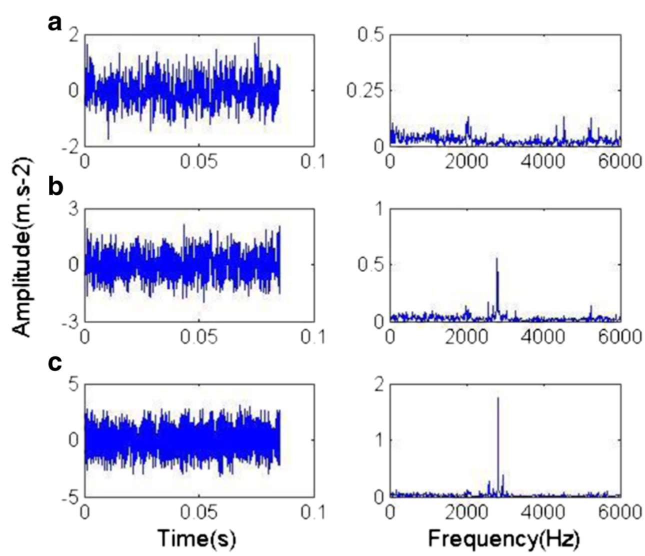

Fig. 3 The measured vibration signal and their FFT. **a** stable state, **b** transition state, and **c** chatter state

the focus of attention in this paper. Early chatter recognition is the recognition at the transition stage. Timely chatter is detected of chatter which may effectively avoid the unfavorable effect on the workpiece and the tools.

2 Feature extraction of chatter

The vibration signals with rich information can be used to detect chatter in the milling process. There are several time-domain and frequency-domain characteristic parameters used in chatter identification. From the thesis [20, 22, 28], this paper selected ten time-domain and four frequency-domain feature parameters, which form the original feature set of chatter.

Supposed $x_i(i = 1, 2, \cdots, N)$ is a signal series, N is the number of data points.

Time-domain feature parameters are described as follows:

(1) (Mean)$a1 = x_m = \frac{1}{n}\sum_{i=1}^{n} x_i$

(2) Standard deviation, $a2 = x_{std} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x_m)^2}$

(3) Root mean square, $a3 = x_{rms} = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(x_i)^2}$

(4) (Peak) $a4 = x_p = \max(|x_i|)$

(5) (Skewness) $a5 = x_{ske} = \frac{\sum_{i=1}^{n}(x_i - x_m)^3}{(n-1)x_{std}^3}$

(6) (Kurtosis) $a6 = x_{kur} = \frac{\sum_{i=1}^{n}(x_i - x_m)^4}{(n-1)x_{std}^4}$

(7) Crest factor, $a7 = CF = \frac{x_p}{x_{rms}}$

(8) Clearance factor, $a8 = CLF = \frac{x_p}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}|x_i|}\right)^2}$

(9) Shape factor, $a9 = SF = \frac{x_{rms}}{\frac{1}{n}\sum_{i=1}^{n}|x_i|}$

(10) Impulse factor, $a10 = IF = \frac{x_p}{\frac{1}{n}\sum_{i=1}^{n}|x_i|}$

When chatter occurs in the machining process, the amplitude and distribution of the time-domain signal may be different from those of the time-domain signal under stable condition. In the time domain, x_m, x_p, and x_{rms} reflect the amplitude and energy of signal. While x_{std}, x_{ske}, x_{kur}, CF, CLF, SF, and IF represent the time series' distribution of signal. Besides, the amplitude and distribution of frequency may change in the frequency.

Frequency-domain feature parameters can be written as:

(11) Mean square frequency, $MSF = \frac{\sum_{j=1}^{m} {f_j}^2 S(f_j)}{\sum_{j=1}^{m} S(f_j)}$

(12) One-step autocorrelation function, $\rho = \frac{\sum_{j=1}^{n}\cos(2\pi f_j \Delta t) S(f_j)}{\sum_{j=1}^{n} S(f_j)}$

(13) Frequency center, $FC = \frac{\sum_{j=1}^{n} f_j S(f_j)}{\sum_{j=1}^{n} S(f_j)}$

(14) Standard frequency, $FV = \frac{\sum_{j=1}^{m}(f_j - FC)^2 S(f_j)}{\sum_{j=1}^{m} S(f_j)}$

In the formula above, $f_j(j = 1, 2, \cdots, m)$ represents the jth frequency in the power spectrum.$S(f_j)$ represents the amplitude of the power spectrum calculated by FFT. Δt represents sample interval. The MSF represents the energy of the vibration signal in the frequency domain. ρ and FC are used to reflect the position change of the main band of the signal. FV is used to describe the signal energy dispersion and concentration in the frequency domain.

However, calculating these four frequency-domain feature parameters takes a lot of time to compute the fast Fourier transform (FFT), so these methods are not suitable for flutter online monitoring. In this study, the fast calculation criterion of frequency-domain characteristic parameters is proposed [31], and the frequency-domain characteristic parameters can be rewritten as follows:

(15) Mean square frequency, $a11 = MSF = \frac{\sum_{i=2}^{n} x^2}{4\pi^2 \sum_{i=1}^{n} x_i^2}$

(16) One-step autocorrelation function, $a12 = \rho = \frac{\sum_{i=2}^{n} x_i x_{i-1}}{\sum_{i=1}^{n} x_i^2}$

(17) Frequency center, $a13 = FC = \frac{\sum_{i=2}^{n} x_i x}{4\pi^2 \sum_{i=1}^{n} x_i^2}$

(18) Standard frequency, $a14 = FV = MSF - 4\pi^2 FC^2$

As mentioned above, some of the feature parameters based on the previous publications have been demonstrated not effective. While, based on different applications, different feature parameters are used to detect the chatter by different researchers. One of the purpose of this paper is to utilize the feature selection method to automatically select the effective chatter identification parameters, rather than depending on human experience.

However, the measured vibration signal generally contains background noise, which is the disadvantage of identifying the chatter both in time domain and frequency domain. Therefore, it is very critical to suppress or eliminate the noise for the feature extraction of chatter. Considering the noise is broadband, the measured signal is decomposed into some narrow band components so that the energy of the noise is dispersed in these narrow bands. Based on the reference [32], wavelet packet transform is the most suitable choice. Wavelet packet transform (WPT) was applied to preprocess the measured vibration signal before feature extraction of chatter. The vibration signal via WPT may be allocated in a specific frequency band. This allows for increasing the signal-to-noise ratio and increasing the sensitivity of chatter features.

3 Wavelet packet transform

Wavelet packet transform is based on wavelet analysis. Not only it can decompose the low frequency component, but also can decompose the high frequency component. It can be a multi-level signal band division in the whole band, improving the time-frequency domain resolution. Therefore, the band signal contains rich information of the original signal.

Assuming that the subspace U_j^n is the closure space of the function $u_n(t)$, and U_j^{2n} is the closure space of the function $u_{2n}(t)$. So the formula $u_n(t)$ that satisfies the two-scale equation is described as:

$$\begin{cases} u_{2n}(t) = \sqrt{2}\sum\limits_{k\in Z} h(k)u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2}\sum\limits_{k\in Z} g(k)u_n(2t-k) \end{cases}$$

where, h_k and g_k are the filter coefficients of the orthogonal wavelet basis. The two coefficients are the orthogonal relationship, and$g(k) = (-1)^k h(1-k)$.

The sequence $\{u_n(t)\}_{n\in Z}$ is constructed by the Eq. (1), which is called the orthogonal wavelet packet determined by the basis function $u_0(t) = \phi(t)$. When $n = 0$, $u_0(t)$ and $u_1(t)$ are orthonormal scaling function $\psi(t)$ and wavelet basis function $\phi(t)$, respectively. The above formula becomes:

$$\begin{cases} \phi(t) = \sqrt{2}\sum\limits_{k\in Z} h(k)\phi(2t-k) \\ \psi(t) = \sqrt{2}\sum\limits_{k\in Z} g(k)\psi(2t-k) \end{cases} \quad (1)$$

The wavelet packets decomposition algorithm:

$$\begin{cases} d_l^{j+1,2n} = \sum\limits_k h_{k-2l} d_k^{j,n} \\ d_l^{j+1,2n+1} = \sum\limits_k g_{k-2l} d_k^{j,n} \end{cases} \quad (2)$$

The wavelet packet reconstruction algorithm:

$$Q_l^{j,n} = \sum_k \left[h_{l-2k} d_l^{j+1,2n} + g_{l-2k} d_l^{j+1,2n+1} \right] \quad (3)$$

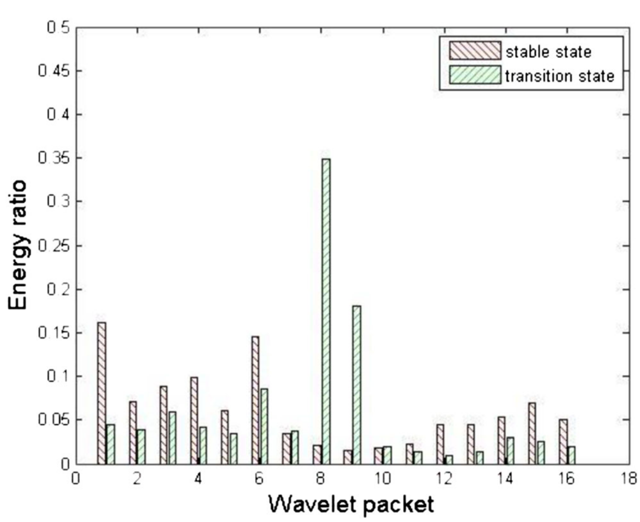

Fig. 4 Comparison of the wavelet packet energy ratio in stable state and transition state

In this paper, the db10 is used as the wavelet basis function which has a better orthogonality. For the stable and transition state, the measured vibration signal was decomposed four levels by WPT. 16 wavelet packets $\left(d_4^i, i = 1, 2, \cdots, 16\right)$ were obtained correspondingly. The energy distribution of the wavelet packet is shown in Fig. 4. When milling is in stable state, energy is mainly distributed at low frequency, and consumption of its energy is mainly used in milling process. When milling is in the transition state, its high frequency components are suppressed, and energy is mainly concentrated in the wavelet packet nodes 8 and 9, where the vibration energy was concentrated around the chatter frequency. The energy and amplitude of vibration signal increase sharply. The wavelet packet nodes 8 and 9 which contain the chatter frequency were selected as the characteristic wavelet packets and reconstructed by the Eq. (3). Then the reconstructed signal $Q(t)$ was obtained. As it is shown in Fig. 5, it is obvious that the reconstructed signals has apparent changes in the early chatter.

Then, the aforementioned ten time-domain and four frequency-domain feature parameters were obtained via calculating the reconstructed signal $Q(t)$. These feature parameters was denoted as the original feature set T = [a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 a12 a13 a14]. The aim was to obtain the more efficient features to monitoring the milling process. In order to select sensitive features of chatter rather than individual experience, the selection feature method of SVM-RFE was applied in the following paper.

4 Feature selection of chatter identification based on SVM-REF

4.1 Theory of the SVM

Support vector machine (SVM) is derived from the thought of the optimal classification surface under the linear separable

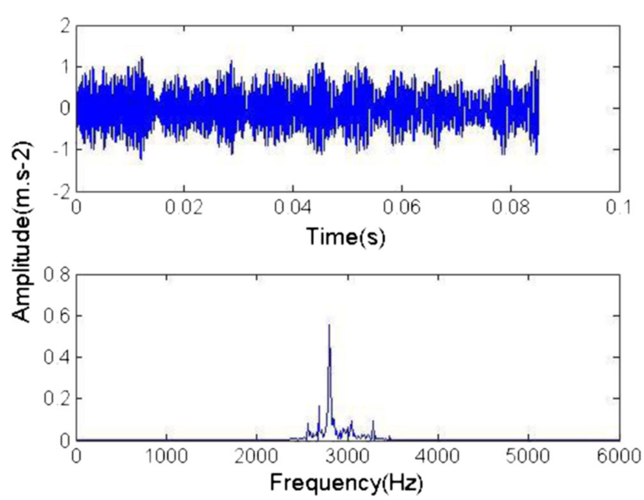

Fig. 5 The reconstructed signal and its spectrum in transition state

case. Support vector regression machine and support vector machine for classification are mainly used. This paper briefly introduces the application of class support vector machine [23, 33].

For the two-class classification problem, supposing there is a training data $\{(x_i, y_i), i = 1, 2, \cdots, n\}$, where n is the number of samples and x_i is n dimensional feature vectors with class labels$y_i \in \{-1, +1\}$.

The goal of SVM is to find an optimist separating hyperplane. Maximizing the margin of the hyperplane is then equivalent to maximizing the distance between the class boundaries. The distance between the margin and the hyperplane is defined as $D = 1/\|\omega\|$. Since the distance are symmetry with respect to hyperplane, the distance between the two margins becomes $2/\|\omega\|^2$. So the problem of the biggest classification margin distance is transformed into finding the minimum $\|\omega\|$. The decision function of SVM is:

$$f(x_i) = \omega \cdot x_i + b \tag{4}$$

where ω is a weight vector and b is a bia.

The position of the separating hyperplane is defined by ω and b. The optimal hyperplane is found out by solving the following constrained optimization problem:

$$\begin{cases} \min L(\omega, b, \xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \\ s.t.\ \ y_i(\omega \cdot \varphi(x_i) + b) - 1 + \xi_i \geq 0, i = 1, 2, \cdots, n \end{cases} \tag{5}$$

where C is the penalty parameter which represents a trade-off between training error and the margin, ξ_i are slack variables, and$\xi_i > 0$.

Using the Lagrange function, the problem can be rewritten as follows:

$$\begin{cases} \min L = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \\ s.t.\ \ \sum_{i=1}^{n} y_i\alpha_i = 0, 0 \leq \alpha_i \leq C,\ \ i = 1, 2, \cdots, n \end{cases} \tag{6}$$

where α_i are called Lagrange multipliers which are the constants and are determined in the optimization process. $K(x_i, x_j)$ is a symmetric and positive kernel function which denotes as $K(x_i, x_j) = \langle\varphi(x_i) \cdot \varphi(x_j)\rangle$, with satisfying Mercer's theory. The derived training algorithm is guaranteed for minimization.

Then, the non-linear decision function of SVM is described as following:

$$\begin{aligned} f(x) &= sign([\omega \cdot \varphi(x)] + b) \\ &= sign\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i, x_j) + b\right) \end{aligned} \tag{7}$$

This decision function is the so-called as SVM. In the practical application, the most commonly adopted kernel function is the radial basis function, which has the form as$K(x_i, x_j) = \exp(-|x_i - x_j|^2/2\sigma^2)$.

4.2 Feature selection of chatter based on SVM-RFE

Support vector machine recursive feature elimination (SVM-RFE) is a feature selection method proposed by Guyon [23, 27]. It is a highly efficient feature ranking criterion based on the weight of SVM classifier. The larger the weight of feature, the greater it impacts on the classification decision. In each iteration of recursive feature elimination (RFE), a trained SVM model is obtained. The feature with the smallest weight is eliminated according to the effect on classification. The remaining features are used to train the SVM model in the next iteration. The iteration process is repeated until all the features have been eliminated. Finally a feature ranking list is obtained. A number of nested feature subsets may be defined to train SVM based on the feature ranking list, then the optimal feature subset is obtained based on the prediction accuracy of SVM classification.

For classification problems, Kohavi [34] indicated that the weight of SVM may be replaced by a cost function. The features are sorted based on the value of the cost function. By calculating the weights, the cost function of each weight could be obtained. The weight vector of SVM classifier could be obtained by Eqs.. (6) and (7):

$$\omega = \sum_i \alpha_i y_i x_i \tag{8}$$

where ω is represented as $\omega = [\omega_1, \omega_2, \cdots, \omega_n]^T$ and n is the number of features in the feature ranking list.

For linear SVM, the cost function of the hth feature is defined by:

$$DJ(h) = (\omega_h)^2 \tag{9}$$

For non-linear SVM, the cost function of the hth feature is described by:

$$DJ(h) = \frac{1}{2}\alpha^T\mathbf{H}\alpha - \frac{1}{2}\alpha^T\mathbf{H}(-h)\alpha \tag{10}$$

where $DJ(h)$ is denoted as the cost function, and h is denoted as the hth removed feature. $\mathbf{H}$ is a matrix with elements $y_i y_i K(x_i, x_j)$, and $\mathbf{H}(-h)$ is a matrix similar $\mathbf{H}$, which removes the hth feature.

Since chatter phenomenon is linked to the dynamic behavior of the machine-tool-workpiece system, the measured vibration signal in milling process is non-linear and non-stationary. So the non-linear cost function is applied in the study. Based on the above analysis, the SVM-RFE feature selection criteria [23] is as follows:

Step 1: Inputs: training examples $X_0 = [x_1, x_2, \cdots, x_i, \cdots, x_n]$ and class labels$y = [y_1, y_2, \cdots, y_i, \cdots, y_n]$;

Step 2: Initialize: subset of surviving features $S = [1, 2, \cdots, k]$, the feature ranking list$r = [\ ]$;

Step 3: The feature ranking process:

(1) Restrict training examples to good feature indices $X = X_0[:, S]$;

(2) Train the classifier $\alpha = SVM - train(X, y)$;

(3) Compute the ranking criteria: $ranking\ DJ(h) = \frac{1}{2}\alpha^T \mathbf{H}\alpha - \frac{1}{2}\alpha^T \mathbf{H}(-h)$, for all h. Find the feature with smallest ranking criterion $f = \arg\min(ranking)$;

(4) Update feature ranked list $r = [S(f), r]$ and eliminate the feature with smallest ranking criterion $S = S(1 : f - 1, f + 1, length(S))$;

(5) Repeat until $S = [\ ]$;

Step 4: Step 4: Output: the feature ranked list$r$.

5 Early scheme of chatter detection

In order to improve the accuracy of chatter identification, the measured vibration signal is preprocessed by a four-level wavelet packet transform. Some wavelet packets with rich chatter information are selected and reconstructed. The feature parameters of ten time domain and four frequency domain are extracted, which form the original feature set. The SVM-RFE criterion is used to select the most sensitive chatter features from the original feature set, which can effectively remove the irrelevant or redundant features. Finally the selected best features are fed into the SVM for pregnant chatter identification. The main scheme of early chatter detection is as follows:

Step 1: Obtain the vibration signal for the milling. The vibration signal of the spindle is collected by the accelerometer.

Step 2: Analysis the vibration signal by WPT. The vibration signal is processed by four-level WPT, and the wavelet packet 8, 9 is selected and reconstructed.

Step 3: Extraction of time-domain and frequency-domain feature. The chatter feature parameters of time domain and frequency domain were calculated via the reconstructed signal, and the 14-dimensional original feature set T is obtained.

Step 4: Select the optimal feature subset based on SVM-RFE. The feature selection method of SVM-RFE was implemented to obtain the feature ranking list. Then a series of feature subset was obtained.

Step 5: Train the SVM classifier. The SVM model is trained based on training data for every feature subset. Correspondingly, the prediction accuracy of testing data is obtained.

Step 6: According to the classification accuracy of chatter, the optimal feature subset is obtained.

6 Chatter identification of the milling

6.1 Feature extraction of chatter

In this paper, identifying the transition from stable to chatter in end milling process was mainly studied. So the vibration signals in stable state and transition state were collected in the milling process. According to the experiment condition Table 1, 60 samples were obtained, where 30 samples were collected in cutting stable state and the others were collected in chatter transition state. Each sample has 1024 data points, which were processed with four-level WPT and 16 time-frequency domain wavelet packets was obtained. In terms of feature extraction, according to the analysis in Section 2, 8, and 9, wavelet packets 8 and 9 were selected and reconstructed. Then, for each sample, ten time-domain feature parameters and four frequency-domain feature parameters were calculated, and 14 feature parameters were obtained, which constructed the original feature set $T = [a1\ a2 \cdots a14\ ]$.

In order to verify the proposed approach in this study, randomly 20 samples were selected as training data from stable and transition samples, respectively. The remained samples were selected as testing data.

6.2 Chatter feature selection based on SVM-RFE

The original feature set may not directly recognize the pregnant chatter state (i.e., the transition state) because of chaos and overlap in structure. This phenomenon is showed in

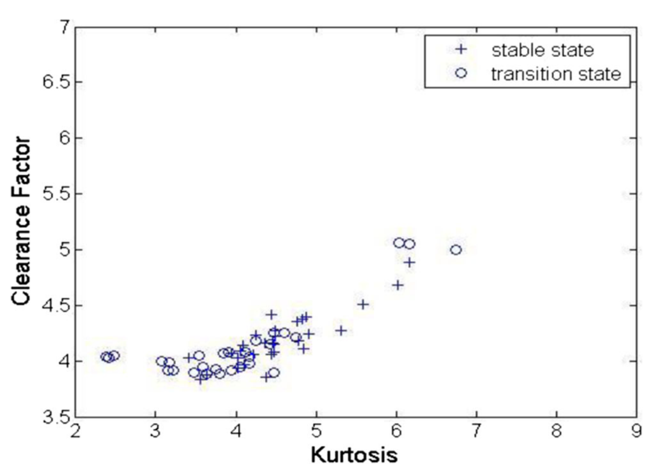

Fig. 6 Spatial distribution of [x_{kur} CLF] between stable state and transition

Table 2 the partial feature subset

Number of used features	Selected feature	Accuracy	Execution time
1	$a12$	80	12.6 ms
2	$a12$, $a10$	100	12.6 ms
3	$a12$, $a10$, $a4$	95	17.4 ms
4	$a12$, $a10$, $a4$, $a7$	80	18.9 ms
…	…	…	…
14	$a12$, $a10$, $a4$, $a7$, $a8$, $a11$, $a5$, $a2$, $a3$, $a9$, $a13$, $a13$, $a6$, $a14$	95	117.3 ms

Fig. 6, which was kurtosis x_{kur} and clearance factor CLF of time-domain parameters from the original feature set. From the picture, it was not effectively in chatter recognition because the selected chatter features exists great relevant. Therefore, the original feature could not be directly fed into the SVM classifier for chatter identification.

The step of feature selection is crucial in order to find which features are the most significant and reliable for chatter detection. To avoid this disadvantage phenomenon, the method of SVM-RFE was implemented to select the best chatter features from the original feature set. Then the redundant or irrelevant features could be eliminated. Features with rich information were selected, chatter identification accuracy of SVM classifier could be improved and the computation time was reduced.

In the algorithm of SVM-RFE, the calculated cost function is realized by using the Lagrange multipliers which is linked to the penalty function C and Gaussian kernel parameter of SVM classifier. As a result, the selection of SVM parameters will directly affect the performance in the process of feature selection. But it is difficult to choose appropriate penalty function C and kernel parameter of SVM classifier, the particle swarm optimization (PSO) [35] was used to optimize the SVM parameters in this paper.

Through the application of particle swarm optimization, the optimal regularization parameter and Gaussian kernel function are $C = 75.68$ and $\gamma = 0.93$, respectively. Based on the algorithm of SVM-RFE, the feature ranking list was obtained, which was represented as [$a12$, $a10$, $a4$, $a7$, $a8$, $a11$, $a5$, $a2$, $a3$, $a9$, $a13$, $a13$, $a6$, $a14$]. The top feature of the feature ranking list had the greatest effect on classification, while the last feature of the feature ranking list had the least influence on classification. Then a series of feature subset was obtained, which is shown in Table 2. And the corresponding prediction accuracy of testing data is obtained. From Table 2, the feature subset comprised of one-step autocorrelation function $a12$ and impulse factor $a11$ was the best, which was selected as the feature vectors of chatter identification. Then the selected two features were fed into SVM to train the SVM model based on the training data. The prediction of classification accuracy was 100% according to the testing data. While, when the feature number was 14, the prediction of classification accuracy decreased to 95%, and the training time and testing time would increase correspondently. It was demonstrated that the selected best features through SVM-RFE could obviously improve the chatter identification accuracy of SVM classification.

6.3 Online verification of chatter detection

The purpose of this section is to verify the selected features comprised of one-step autocorrelation function $a12$ and impulse factor $a10$ in online detection of chatter onset. From Fig. 7, it was seen that there was no obvious change of the vibration signal amplitude before the time is 5.1 s, as shown in line 2 in Fig. 7. While the increasing trend was emerged through the selected two feature parameters, which means the early evidence of chatter was occurred at 4.2 s during the milling process, as shown in line 1 in Fig. 7. If suppression chatter could be carried out at this moment, there would be enough time to avoid the disadvantageous effect.

From Fig. 7, it was also indicated that the one-step autocorrelation function $a12$ reflected the change of dominant

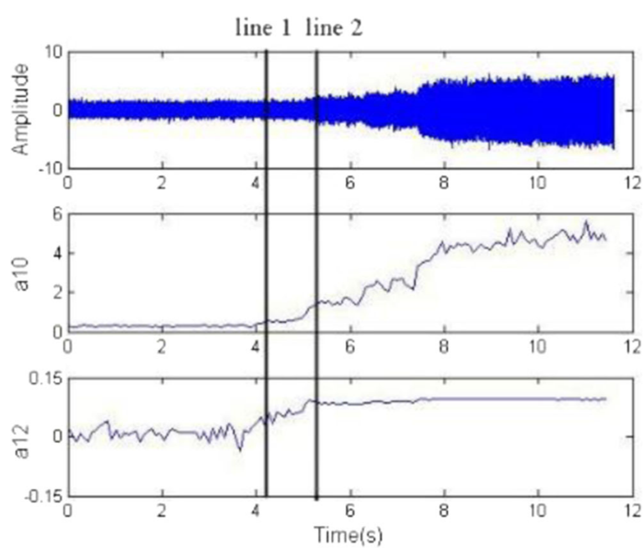

Fig. 7 The selected features $a10$ and $a12$ in chatter development process

frequency band in frequency domain, gathered around the chatter frequencies with the increase of chatter intensity when chatter occurred. And the impulse factor reflected the frequency of the vibration pulses, a dimensionless feature parameter on time domain, represented the increasing degree of collision between the milling tool and workpiece as time went on. Hence, the selected two feature parameters based on SVM-RFE fully reflected the chatter phenomenon from both time domain and frequency domain during the milling process.

7 Conclusion

This paper proposed a novel method of chatter detection in milling machines based on wavelet packet transform and support vector machine recursive feature elimination, which overcame the influence of artificial experience of selecting the chatter identification feature parameters. The collected vibration signal was preprocessed by wavelet packet transform regarded as a filter, which allowed for increasing the signal-to-noise ratio and the sensitive features. Features were ranked by the value of cost function via SVM-RFE. Through the feature ranking list, the optimal feature subset was obtained, which was comprised of one-step autocorrelation function $a12$ and impulse factor $a10$. The experimental results showed that the proposed technique offered good chatter detection based on the two selected chatter features, which not only improve the chatter identification accuracy but also reduce the computational time.

The proposed method is quite encouraging. The sensitive chatter features are selected based on the algorithm of SVM-RFE rather than individual experience, and the good classification accuracy guarantees its reliability only under the specific conditions. However, when the cutting condition changes, the natural frequency will change accordingly. In order to solve this issue, a simple hammer test is implemented to obtain the system natural frequency, and the wavelet packets around the twice natural frequency of system are selected. If that was done, the proposed method in this paper may be applied in other condition. In the future work, since industrial applications are much more complicated, more experimental conditions including feed rate should be carried out in more complicated milling conditions.

Funding information This project was supported by National Key Research and Development Program of China (Grant No. 2017YFB1104600).

References

1. Quintana G, Ciurana J (2011) Chatter in machining processes: a review. Int J Mach Tools Manuf 51(5):363–376
2. Altinas Y, Weck M (2004) Chatter stability of metal cutting and grinding. Ann CIRP 53(2):619–642
3. Siddhpura M, Paurobally R (2012) A review of chatter vibration research in turning. Int J Mach Tools Manuf 61(10):27–47
4. Li XQ, Wong YS, Nee AYC (1997) Tool wear and chatter detection using the coherence function of two crossed accelerations. Int J Mach Tools Manuf 37(4):425–435
5. Hutton DV, Hu F (1999) AE monitoring of tool wear in end milling using time domain averaging. J Manuf Sci Eng 121:8–12
6. Soliman E, Ismail F (1997) Chatter detection by monitoring spindle drive current. Int J Adv Manuf Technol 13(1):27–34
7. Quintana G, Ciurana J, Teixidor D (2008) A new experimental methodology for identification of stability lobes diagram in milling operations. Int J Mach Tools Manuf 48(15):1637–1645
8. Tangjitsitcharoen S, Saksri T, Ratanakuakangwan S (2013) Advance in chatter detection in ball end milling process by utilizing wavelet transform. J Intell Manuf 26(3):485–499
9. Lamraoui M, Thomas M, El Badaoui M, Girardin F (2014) Indicators for monitoring chatter in milling based on instantaneous angular speeds. Mech Syst Signal Process 44(1–2):72–85
10. Kuljanic E, Sortino M, Totis G (2008) Multisensor approaches for chatter detection in milling. J Sound Vib 312(4–5):672–693
11. Jia G (2015) Research on chatter and tool wear recognition methods in the difficult-to-cut materials milling. Huazhong University of Science and Technology, Wuhan
12. Somkiat T, Narongsak P (2013) Development of chatter detection in milling processes. Int J Adv Manuf Technol 65(5):919–927
13. Somkiat T (2012) Analysis of chatter in ball end milling by wavelet transform. World Acad Sci Eng Technol 71:1075–1081
14. Somkiat T, Tanintorn S, Suthas R (2015) Advance in chatter detection in ball end milling process by utilizing wavelet transform. J Intell Manuf 26:485–499
15. Sun Y, Zhuang C, Xiong Z (2014) Real-time chatter detection using the weighted wavelet packet entropy, IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp 1652–1657
16. Yao Z, Mei D, Chen Z (2010) On-line chatter detection and identification based on wavelet and support vector machine. J Mater Process Technol 210(5):713–719
17. Chen B, Yang J, Zhao J, Ren J (2015) Milling chatter prediction based on the information entropy and support vector machine, International Industrial Informatics and Computer Engineering Conference (IIICEC), pp 376–380
18. Lamraoui M, El Badaoui M, Guillet F (2014) Chatter detection in CNC milling processes based on Wiener-SVM approach and using only motor current signals. Mech Mach Sci 23:567–578
19. Lamraoui M, Barakat M, Thomas M, El Badaoui M (2015) Chatter detection in milling machines by neural network classification and feature selection. J Vib Control 2(7):1251–1266
20. Lei Y, He Z, Zi Y, Chen X (2008) New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. Mech Syst Signal Process 22(2):419–435
21. Li S, Zheng P-l (2012) Hydraulic system fault diagnosis method based on GA-PLS and AdaBoost. J Appl Stat Manag 31(1):26–31
22. Li W, Xu Y (2010) Gearbox incipient fault diagnosis using feature sample selection and principal component analysis. Int J Model Identif Control 10(3–4):246–254
23. Guyon I, Weston J, Barnhill S (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422
24. Miranda J, Montoya R, Weber R (2005) Linear penalization support vector machines for feature selection. In: Pal SK et al (eds) PReMI 2005, LNCS, vol 3776. Springer-Verlag, Berlin, pp 188–192
25. Tang Y, Zhang YQ, Huang Z (2007) Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis, computational biology and bioinformatics. Trans IEEE/ACM 4(3):365–381

26. Hidalgo-Muñoz AR, López MM et al (2013) Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing. Expert Syst Appl 40(6):2102–2108
27. Qian S, Sun Y, Xiong Z (2015) Intelligent chatter detection based on wavelet packet node energy and LSSVM-RFE, Advanced Intelligent Mechatronics (AIM), 2015 I.E. International Conference on, pp 1514–1519
28. Altintas Y, Budak E (1995) Analytical prediction of stability lobes in milling. Ann CIRP 44(1):357–362
29. Suh CS, Khurjekar PP, Yang B (2002) Characterisation and identification of dynamic instability in milling operation. Mech Syst Signal Process 16(5):853–872
30. Zatarain M, Muñoa J, Peigné G, Insperger T (2006) Analysis of the influence of mill helix angle on chatter stability. Ann CIRP 55(1):365–368
31. Shi HM (2003) Metal cutting theory and practice: a new perspective (in Chinese). Huazhong University of Science and Technology Press, Wuhan
32. Hongrui C, Yaguo L, He Z (2013) Chatter identification in end milling process using wavelet packets and Hilbert-Huang transform. Int J Mach Tools Manuf 69(3):11–19
33. Vapnik V (1998) Statistical learning theory. Wiley
34. Kohavi R, John G (1997) Wrappers for feature subset selection. Artil Intell 12:273–324
35. Thelaidjia T, Chenikher S (2014) A new approach of preprocessing with SVM optimization based on PSO for bearing fault diagnosis, International Conference on Hybrid Intelligent Systems, pp 319–324