MULTI-CATEGORIES TOOL WEAR CLASSIFICATION IN MICRO-MILLING

ZHU KUNPENG

(*M. Eng, Huazhong Univ. of Sci. & Tech.*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgements

I am particularly grateful to my supervisors, Associate Professor G. S. Hong and Associate Professor Y. S. Wong, for motivating and directing me in this project throughout the years. I want to thank them for their motivation, support, and critique about the work. Professor Hong and Prof. Wong always offer me wise and constructive feedback and advice during the fortnightly meetings. The supervisor's depth of knowledge, insight and untiring work ethic has been and will continue to be a source of inspiration to me.

Thanks to Mr. Simon Tan, Mr. S. C. Lim, Mr. C. L. Wong, and all the technicians at Advanced Manufacturing Lab of NUS for their kind and quick technical assistance during my experiments.

I have also benefited from discussion with many of my colleague. In particular, I would like to thank Dr. Wang Zhigang, Dr. Wang Wenhui, Mr. Dong Jianfei, and many others, for their enlightening discussion and suggestions. More particularly, I would like to thank those who provided help in difficult times.

I also would like to thank National University of Singapore for offering me research scholarship and excellent research facilities. The abundant professional books in the NUS library are also to my benefit.

Finally, I would like to devote this thesis to my family for their love and understanding.

Table of Contents

Summary

In-process monitoring of tool conditions in micro-machining can significantly improve machining efficiency, and minimize inaccuracy and occurrence of tool breakage due to the high tool wear rate and high precision requirement associated with the dimensions to be produced at micro-level. Tool condition monitoring in micro machining poses new challenges compared to conventional machining. In this thesis, a multi-category classification approach for tool flank wear state identification in micro-milling is proposed. For this purpose, three issues are discussed and addressed.

The first concerns force denoising. Force has been found to be most sensitive in tool condition monitoring. In micro-milling, the comparatively small cutting force signal is prone to contamination by relatively large noises, and as a result it is important to denoise the force signal before further processing it. However, the traditional denoising methods, based on Gaussian noise assumption, is not as effective in this situation because the noise is found to contain high non-Gaussian component in the experiment. An approach has been developed that employs fixed-point independent component analysis (FastICA). It assumes that the noise is the source and force signal to be instantaneous mixtures of sources, and treats the signal denoising as a blind source separation (BSS) approach. The results show that FastICA effectively separates both Gaussian and non-Gaussian noise sources, which is needed in the study.

The second issue concerns feature dimension reduction. Numerous features based on the force signal contain redundant information or are less sensitive to tool state discrimination. These features can be eliminated for reduced computation and more robust modeling. Fisher's linear discriminant analysis (FDA) is adapted for this purpose. In the discriminant selection, features are chosen to maximize class separation and are ranked by their separation ability between different classes. Other popular feature dimension reduction methods, such as principal component analysis (PCA) and automatic relevance determination (ARD), are also discussed and compared with the

discriminant method with their classification rate. The reasons that the FDA are superior to both PCA and ARD in feature selection are also discussed.

The third issue concerns tool wear state estimation. The existing approaches have been found not to be suitable for tool wear monitoring in micro milling. Continuous Hidden Markov models (HMMs) are adapted for stochastic modeling of the tool wear in micro-milling, and estimation of the tool wear state based on the cutting force features. A detailed study on the selection of HMM structures for tool condition monitoring (TCM) is presented. In the framework of HMMs, the most discriminant features of the cutting forces are selected from both time and wavelet domain first, and then these features are iteratively learned by the HMMs to map the relationships between force features and tool wear states. Different tool wear states are modeled as separate HMMs. The tool wear state is then classified by the HMM that has the maximum probability to indicate the test features. Experimental studies on the tool state estimation in the micro-milling of pure copper and steel illustrate the effectiveness and potential of these methods.

List of Tables

List of Figures

List of Symbols

TCM:	Tool Condition Monitoring
AR:	Autoregressive
ARMA:	Autoregressive Moving Average
PCA:	Principal Component Analysis
LDA:	Linear Discriminant Analysis
NN:	Neural Networks
MLP	Multilayer Perceptron,
SOM	Self-Organising Map
RNN:	Recurrent Neural Network
ART:	Adaptive Resonance Theory
SVM:	Support Vector Machines
GMM:	Gaussian Mixture Models
HMM:	Hidden Markov Models
AE:	Acoustic Emission
DWT:	Discrete Wavelet Transform
MAP:	Maximum a Posteriori
PSD:	Power Spectrum Density
DFT:	Discrete Fourier Transform
FFT:	Fast Fourier Transform
STFT:	Short-Time Fourier Transform
LTI:	Linear Time Invariant System
WT:	Wavelet Analysis
CWT:	Continuous Wavelet Analysis
DWT:	Discrete Wavelet Analysis
FWT:	Fast Wavelet Transform
MRA:	Multi-Resolution Analysis
CQF:	Conjugate Quadratic Filters
WPD:	Wavelet Packet Decomposition
WVD:	Wigner-Ville Distribution
CWD:	Choi-Williams Distribution

KLT:	Karhunen-Loeve Transform
HHT:	Hilbert–Huang Transform
EMD:	Empirical Mode Decomposition
SNR:	Signal to Noise Ratio
L.E.:	Lipschitz Exponent
WTMM:	Wavelet Transform Modulus Maxima
Pdf:	Probability Density Function
FDR:	Fisher's Discriminant Ratio
ICA:	Independent Component Analysis
Fastica:	Fixed Point ICA
EM:	Expectation Maximization
AIC:	Akaike Information Criterion
BIC:	Bayesian Information Criteria
CV:	Cross Validation
HMM:	Hidden Markov Model
CHMM:	Continuous Hidden Markov Model
DHMM:	Discrete Hidden Markov Model
VB	Average Wear Land Width
Vb_{max}	Maximum Wear Land Width
KB	Crater Width
KM	Crater Center Distance
KT	Crater Depth
KA	Crater Area
RPM:	Revolution Per Minute
R	Radius of cutting edge
μ:	Mean
σ^2:	Variance
F_X:	Feed force
F_Y:	Radial force
F_Z:	Vertical force
$\Delta F_a(i)$:	Dynamic component of cutting force
$F_{med}(i)$:	Average force component

$x(t)$:	Any sensory signal from machining
$y(t)$:	Extracted features
ω_i:	Class i
$\langle x(t), y(t) \rangle$:	Inner product of two signal $x(t)$ And $y(t)$
$c_{j,k}$:	Wavelet coefficients
$d^n_{j,k}(t)$:	Wavelet package coefficients
$f(t)$:	Any mathematical signal $f(t) \in L^2_R$
$\hat{f}(\omega)$:	Fourier transform of $f(t)$
f_s:	Sampling frequency
T_s:	Sampling Interval
σ_t:	Resolution of time
σ_w:	Resolution of frequency
$\psi(t)$:	Wavelet function
$\varphi(t)$:	Scaling function
$W_{j,n,k}$:	Wavelet packet
$c_{j,k}$:	Scaling coefficient
$d_{j,k}$:	Wavelet coefficient
$d^n_{j,k}$:	Wavelet packet coefficient
H_0:	Low pass filter
H_1:	High pass filter
$\oplus$:	Orthogonal complement
s:	Original source
v_i:	Additive noise
W:	Mixing matrix.
$J(Y)$:	Neg-entropy of Y
f_r :	Eigen frequency
$\hat{*}$:	Estimate of *
S_i:	Covariance matrix of class i

S_W :	Within-class covariance matrix
S_B :	Between-class covariance matrix
E_j :	Average wavelet packet coefficients energy
$s(t)$:	Hidden state
VB:	Flank wear
$v(t)$:	Wear rate
$\varepsilon(t)$:	Noise term
$w(t)$:	Tool wear value at time t
O :	HMM observed output
Q :	HMM state sequence
$\alpha_t(i)$:	Forward probability
$\beta_t(i)$:	Backward probability
$V_t(i)$:	Probability of the most likely state sequence at time t
$\lambda = (\pi, A, B)$:	Parameter sets of general HMM
π	Initial state distribution, the probability of state *i* at time $t = 0$.
$A = [a_{ij}]$	Probability of transiting to state *j* given current state *i*.
$B = [b_{jk}]$	Probability of observing feature *k* given current state *j*.
$N(\mu, \Sigma)$:	Gaussian (normal) distribution with mean μ and covariance matrix Σ
$\lambda = (\pi, A, c, \mu, \Sigma)$:	Parameter sets of continuous gaussian mixture hmm
$\gamma_t(i, j)$:	Probability of being at state S_i at time t and state S_j at time $t+1$
d_i :	the *i-th* distance
v_{ri} :	the *i-th* reference spindle speed
f_{ri} :	the *i-th* reference feed rate
d_{ri} :	the *i-th* reference vertical depth of cut
d_{rri} :	the *i-th* reference radial depth of cut

Chapter 1
Introduction

1.1 Micro-machining

The rapid development of micro mechanical and mechatronic systems requires manufacturing of micro components and features, based on processes such as micro fabrication and micro-machining. Micro fabrication techniques, such as lithography and etching, are limited to certain type of materials (mostly silicon) and features (2 or 2.5 dimensions) to the required quality [210] [53]. Micro-machining involving the use of a cutting tool can overcome these shortcomings [50] [27] [126]. It can machine hard material such as copper [162] and steel [208] with complex shapes in three dimensions, and bridges the gap between the nano-technology and macro-worlds. Micro machining is a precision/ultra-precision machining technology where the tolerances, cutting depth, and even part sizes are in micro scale.

Compared to conventional machining, the size scales of some of the micromachining parameters are largely reduced to meet the requirement of miniature products. Figure 1.1 shows the capability of micromachining relative to other processes such as laser machining, Electrical Discharge Machining (EDM), grinding and the Lithography (LIGA) process. It can be seen that surface roughness Ra values in the range down to almost 5 nm can be attained for features down to 1μm. For the case of conventional machining e.g. CNC turning and milling machines, accuracies of 10- 100 μm can be achieved.

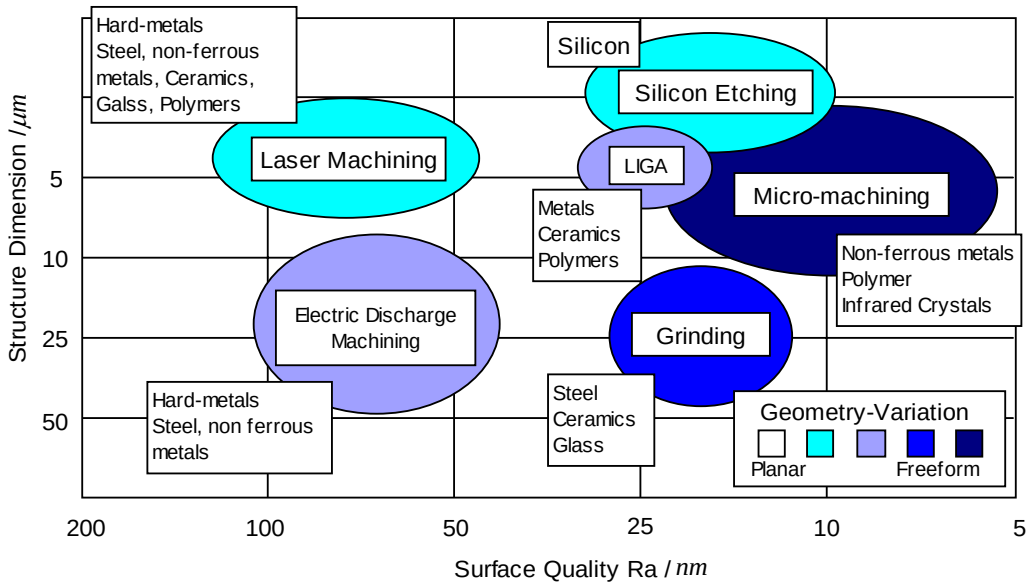


Figure 1.1: Micromachining relative to other machining processes [27]

1.2 Micro-milling and Tool Wear

One very versatile micro-machining process is micro-milling. Micro-milling is a scaled down version of traditional milling with tool diameter of generally between 20 and 800 μm. Peripheral end-milling and slot milling present some of the severe machining environments of micro-machining processes. It has advantages over other micro-machining techniques in the range of workable materials and the free-form 3D micro structures with high aspect ratios and high geometric complexity [30] [58] [207].

One CNC machine used for micro-milling is the Markino V55, as shown in Figure 1.2. It is located at Advanced Manufacturing Laboratory of National University of Singapore, and is used in this project.

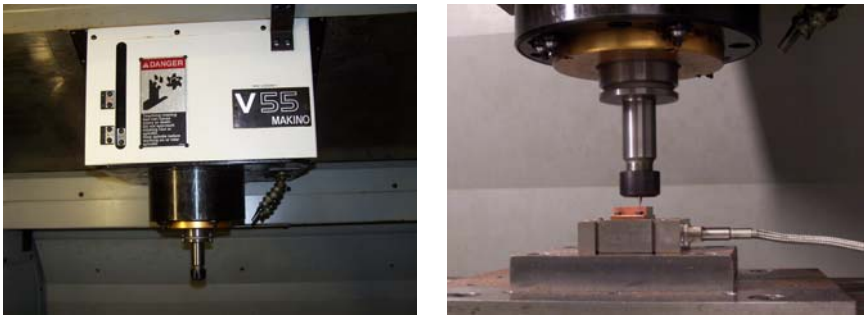
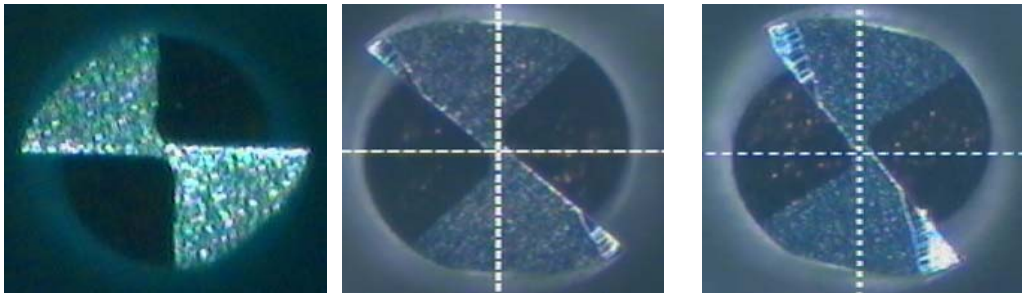


Figure 1.2 Micro-end milling machine

In micro-machining, with the miniaturisation of the cutting tool (<1 mm in diameter), and high speed (>10,000 rpm) is used, the tool wears quickly. Tool wear is defined as the change of shape of the tool from its original shape during cutting, resulting from the gradual loss of tool material. The contact between the cutting tool and the workpiece and chips causes the shape of the tool to change. This tool wear phenomenon (Figure 1.3) has a major influence on machining efficiency, precision of workpiece dimensions, surface roughness, and may lead to product failure. Figure 1.3 shows the focused view of a carbide tool with diameter of 500μm. The modeling of the relationship between different tool wear levels and the corresponding cutting force features is the main focus and broadly discussed in this thesis.



a) Fresh tool b) progressive wear c) accelarated wear

Figure 1.3 Flank face and flank wear of a micro end mill tool with diameter of 500μm

1.3 Problem statement

Tool wear is one of the major limiting factors in high-speed machining [92]. It is critical to monitor the tool wear in micro-machining due to the high precision requirement. Compared to conventional machining, different difficulties are encountered in the identification of tool conditions in micro-machining. For example, chip flow characteristics and vibration are not easily observable in micro-machining, due to minimum chip thickness effect [78] [202] [124], and the very small vibration caused by variations of the cutting force [85] [91]. It is almost impossible to carry out direct (vision) approach under such a high spindle speed. Indirect approach using cutting force is more appropriate for the TCM in micro-machining.

Cutting force has been found to correlate well with tool conditions in machining and most effective as sensor signal for tool wear monitoring [8] [26]. Figure 1.4 shows a

typical sample of the cutting force and noise in our experiments. It can be observed that the force signal is very low and highly non-stationary. The signal-to-noise ratio is relatively much lower than that of conventional machining. Effective signal processing for this highly non-stationary and noisy signal for tool condition monitoring (TCM) is a primary challenge in micro machining [218]. A difficulty of denoising in micro-milling is that the current denoising algorithms are based on Gaussian noise assumption, but the noise in micro-milling contains both Gaussian noise and non-Gaussian component [217] [218]. The noise component in the signal for monitoring micro-machining is usually very high [190] [186] [218] and difficult to separate [190] [218]. It has been found that the noise harmonics distribute widely in frequency domain. The traditional approaches are not effective in this condition.

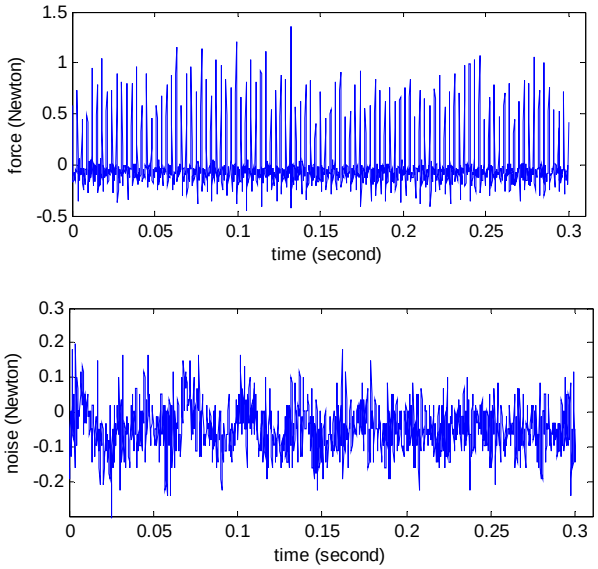


Figure 1.4 Typical cutting force and noise in micro-milling

For the TCM, the drawback of current feature extraction approaches is that it transforms the features into anther space, losing their original physical meanings, which are very important in understanding and explaining the tool states. And some time the extracted features may be less sensitive to tool state discrimination, since the feature extraction is generally an unsupervised approach, and may not include class information

On the other hand, according to ISO 8688, the threshold to determine the tool life is a maximum flank wear of 0.3mm in conventional machining [79]. This has to be

redefined in micro-machining because the total cutting edge is less than of 0.3mm generally. According to Tansel et al [190] [186], and Rahman et al [162], any sense of changes of the cutting edge is wear. From this point of view, it is a matter of different degrees of wear and hence, instead of a single indicator, multi-category identification and classification of tool wear for monitoring progressive wear of the tool is more appropriate. The existing TCM approaches are not suitable for tool wear monitoring in micro milling. The tool wear states have to be redefined generally and estimated in multi-category for the high product precision requirement.

1.4 Objective of this Work

The aim of this thesis is to develop a framework for noise-robust multi-category tool flank wear estimation in micro-milling based on cutting forces. Three problems are to be addressed to achieve this goal.

1. The first is force denoising. This approach intends to develop a general tool for signal denoising when the desired signal is contaminated with non-Gaussian noise. The signal denoising is to be solved as a signal separation approach, since the non-Gaussian noise is typically like a certain kind of signal. It assumes that the noise is source and treating the signal denoising as blind source separation (BSS). In this way, it widens the topics of signal denoising and provides a general tool for signal denoising in case of non-Gaussian noise.

2. The second is feature extraction and selection. To overcome this pitfall of current feature extraction methods, an approach is presented in a framework of how to analyze the cutting force signals both in time and frequency domain, and then select the features by their class-separation scores. The top ranked features are then used for tool state estimation. This can be appropriately adapt to any other machining types, micro-machining or conventional machining, and thus provides a general framework for feature extraction and selection in tool condition monitoring.

3. The last problem is to develop a noise-robust multi-category tool wear state classification system in micro-milling. It is required to effectively process the highly non-stationary and noisy signal in micro-milling, and to provide a localized model of

the tool condition. The proposed approach intends to be noise-robust and not oscillating between different tool states.

1.5 Organization of this thesis

The thesis consists of nine chapters as follows.

Chapter 1 introduces the background and objectives of this thesis.

Chapter 2 reviews TCM literature. As little literature is found in micro machining TCM, the review consists mainly of traditional machining TCM, which provides necessary background to understand the needs for micro machining TCM.

Chapter 3 introduces and reviews wavelet analysis according to their different applications in TCM, both for signal analysis and feature extraction.

Chapter 4 presents a general description of TCM problem from a pattern recognition point of view, and describes the framework to be developed.

Chapter 5 discusses the cutting force denoising with ICA methods. Signal denoising is generally not discussed in details in conventional machining TCM, but is important in micro-machining since the signal to noise ratio (SNR) is very low.

Chapter 6 discusses the wavelet packet decomposition of the cutting forces for feature extraction and then the discriminant feature selection methods for HMMs modeling.

Chapter 7 provides a brief background on HMMs, and the basic idea of HMMs for TCM modeling. The selection and modeling issues of HMMs for micro milling TCM are discussed in details in this section.

Chapter 8 shows the HMM estimation results from all the micro-milling experiments.

Chapter 9 concludes the thesis and discusses the limitations of the approaches proposed and suggestions for future work.

Chapter 2

Literature Review

Tool condition monitoring is extensively studied by many researchers since late 1980s. There is vast literature on tool wear monitoring of conventional machining, but for micro-machining there is significant less available since interests of micro-machining started in late 1990s. However, the monitoring approaches and signal processing algorithms for the conventional machining are suitable valid for micromachining with appropriate adaptations and adjustments. Literature for both conventional machining and micromachining tool monitoring can be viewed according to the three functional stages: measurement, feature extraction, and classification.

2.1 Overview of Tool Condition Monitoring

In tool condition monitoring, the aim is to apply appropriate sensor signal processing and pattern recognition techniques to identify and predict the cutting tool state, so as to reduce loss brought about by tool wear or tool failure. Effective TCM systems can improve productivity and ensure workpiece quality, and hence, has a major influence in machining economics [174] [24]. Tool condition monitoring is extensively studied by many researchers since late 1980s. Many of the reported research works are reviewed in [26], [40], [117].

For successful implementation of tool wear monitoring systems, the sensor signals should be sensitive to tool conditions in the harsh machining environment. Various sensors are amiable to detect the tool wear, either directly from the tool or indirectly from the workpiece or machine table. Since tool wear is typically defined according to the geometrical changes in the tool, direct monitoring methods such as vision and optical approaches, which measure the geometric parameters of the cutting tool, have been developed [205] [147] [98]. The direct methods have advantages of measuring actual geometric changes arising from wear of tool. The dashed line of figure 2.1

illustrated the idea of direct approach: The tool state is observed from images, which is directly measured or after some image processing techniques.

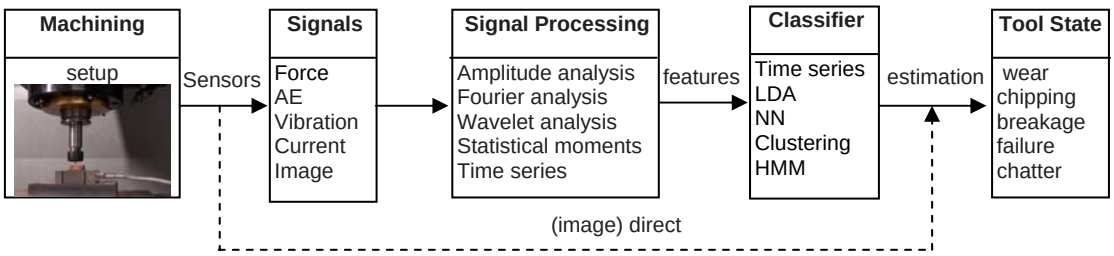


Figure 2.1 Generic structure of TCM

dashed line: direct approach, solid line: indirect approach

However, direct measurements are very difficult to implement because of the continuous contact between the tool and the workpiece, and often the presence of coolant fluids. This difficulty severely limits the application of direct approach, and generally, people seek for indirect methods. The indirect approaches are achieved by correlating or deducing suitable sensor signals to tool wear states. They have the advantages of less complicated setup and suitability for practical application. For indirect approaches, we cannot directly measure the current tool wear, but estimate it from the extracted measurable signal features. These signal features are extracted from several signal processing steps (the solid line of figure 1) for sensitive and robust representation of its corresponding state.

Indirect methods, such as those based on sensing of the cutting forces [58] [207] [92] [78] [202] [124] [85], vibrations [91] [190] [186] [218] [79], acoustic emission [8] [26] [212] [40] [72] [169], and motor/feed current [144] [145] [181] [205] [31] [108] have been the most employed and reported for TCM by many researchers. Detailed works on the design and implementation of these indirect approaches for TCM have been reported in [233] [70] [54]. It is difficult to generalize tool condition monitoring as tool wear is influenced by a number of variables such as tool material, workpiece material, cutting temperature, machining parameters, environmental conditions and lubricants.

2.2 Tool Wear Definition and Tool Wear Mechanism

Tool Wear is defined as the change of shape of the tool from its original shape during cutting, resulting from the gradual loss of tool material. Figure 2.2 (a) shows the generic geometry and wear definition. The high contact stress between the tool rake-face and the chip causes severe friction at the rake face. The result is a variety of wear patterns and blemishes which can be observed at the rake face and the flank face. Figure 2.2 (b) shows the geometry and different measurement criteria, which forms the basis of the following discussion on tool wear.

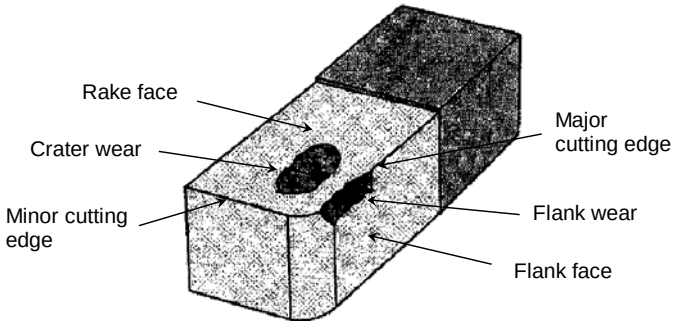


Figure 2.2 (a) Tool geometry

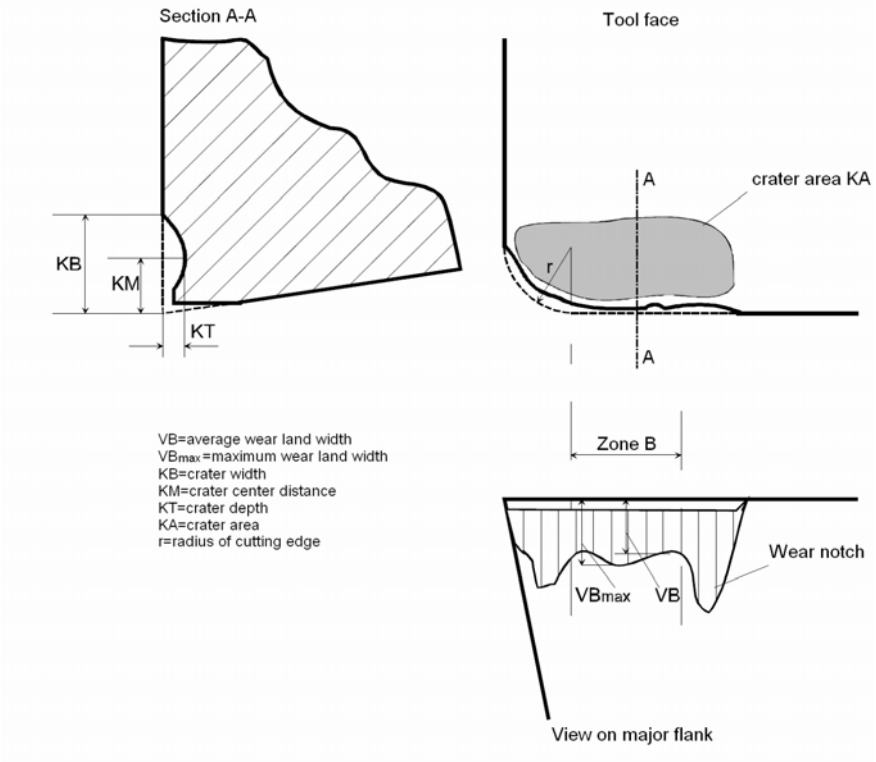


Figure 2.2 (b) tool wear definition

Figure 2.2 Tool geometry and wear definition

Flank wear

Wear on the flank face is called flank wear. Flank wear is the most common wear and results from abrasive wear of the cutting edge against the machined surface. Flank wear is measured by using the average and maximum wear land size VB and VB_{max}. Generally there are three stages of tool flank wear. A typical tool life curve is shown in figure 2.3 with the three regions.

1. Initial wear region:

This is manifested in the form of micro-cracking, surface oxidation and carbon loss layer, as well as micro-roughness at the cutting tool tip in tool manufacturing. For the new cutting edge, the small contact area and high contact pressure will result in high wear rate.

2. Progressive wear region:

After the initial wear (cutting edge rounding), the micro-roughness is improved. In this region the wear size is proportional to the cutting time. The wear rate is relatively constant.

3. Severe wear region:

When the wear size increases to a critical value VB_{max}, the surface roughness of the machined surface decreases, cutting force and temperature increase rapidly, and the wear rate increases exponentially, and then the tool loses its cutting ability. In practice, the tool should be replaced before entering this region.

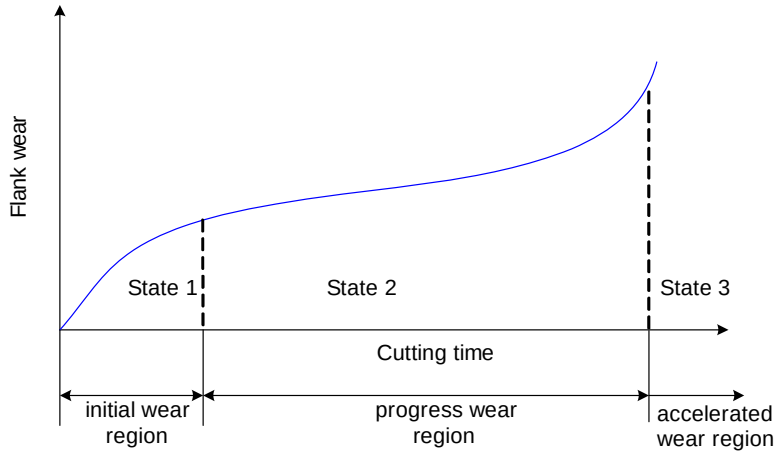


Figure 2.3 Three stages of tool flank wear

Crater wear

The chip flows across the rake face, resulting in severe friction between the chip and rake face, and leaves a scar on the rake face which usually parallels to the major cutting edge. The crater wear commonly appears in high speed machining. The parameters used to measure the crater wear can be seen in the figure 2.2 (b). The crater depth KT is the most commonly used parameter in evaluating the rake face wear.

Notch wear

Notch wear is a combined flank and crater wear that occurs close to the point where the major cutting edge intersects the workpiece surface. It is also common in machining of materials with high work-hardening characteristics, including many stainless steels and heat-resistant nickel or chromium alloys.

Tool failure

Tool failure is the final result of tool wear with complete removal of the cutting point. This may come about by temperature rise, which causes the tool tip to soften until it flows plastically at very low shear stress. Tool failure by melting and plastic flow is most common in carbon and high-speed-steel tools.

Chipping

Chipping involves removal of relatively large discrete particles of tool material. Tools subjected to discontinuous cutting conditions are particularly prone to chipping. Built-up edge formation also has a tendency to promote tool chipping when the material is removed. Edge chipping is common in milling. Chipping may occur when the tool first contacts the part or, more commonly, when it exits the part.

Tool breakage

In milling, tools are subjected to cyclical thermal and mechanical loads. The cyclical variations in temperature in milling induce cyclical thermal stress as the surface layer of the tool expands and contracts. This can lead to the formation of thermal fatigue cracks near the cutting edge. The growth of these cracks eventually leads to edge chipping or tool breakage.

Influence of tool wear on cutting forces

Flank wear, chipping, and crater wear affect the performance of the cutting tool in various ways. The cutting forces are normally increased by wear of the tool. Flank wear and chipping almost invariably increase the cutting forces due to increased rubbing forces. Crater wear may, however, under certain circumstances, reduce forces by effectively increasing the rake angle of the tool.

2.3 Measurement Methods

For successful implementation of a tool wear monitoring system, sensor signal should be reliable and sensitive to tool condition in the harsh manufacturing environment. Since the tool wear is defined according to the geometrical changes in the tool, direct monitoring methods, which measure the geometric parameters of the cutting tool, were developed [98] [147]. The direct methods have advantages of continuous measurement of actual tool wear directly from the tool while a parameter correlated to tool wear needs to be measured in indirect methods to determine the remaining tool life. However, direct measurements are very difficult due to continuous contact between the tool and the workpiece during machining. The indirect methods have the advantages of less complexity and suitability for practical application [26]; thus, indirect methods such as forces, vibrations, acoustic emission have been the most commonly used signals and more successfully achieved for TCM by many researchers.

Force

Force is one of the most important signals in milling process monitoring. It has been reported by researchers that cutting forces contain reliable information on cutting conditions and most effective for tool wear monitoring [21] [22]. Altintas et al [8] developed a model that detects tool breakage by measuring the forces. It was found that the first and the second differences of mean resultant forces between adjacent teeth are quite effective in the recognition of tool breakage. Other studies were also reported from cutting force analysis with time series analysis [9], power spectrum analysis [51], and wavelet analysis [102][211].

Acoustic emission (AE)

The advantage of AE is that the signal measured is a source of engagement where the chip is formed. It was introduced by Diei [39]. Li [112] used the AE signal and electric feed current signal to detect tool breakage using the discrete wavelet transform (DWT). The envelope detection method calculates the second difference of each wavelet coefficient for comparison with the tolerance threshold to predict tool state.

Vibration

The occurrence of vibrations during milling affects the surface finish and tool life. Vibration arises from wrongly chosen machining parameters such as depth of cut and feed per tooth. Huang et al [74] successfully detected tool failure with an accelerometer and showed that it can be implemented and used for milling on line monitoring. It was also studied by Dimla [40], Lee et al [103], and Li et al [114]. Generally, the vibration for TCM is brought by the cutting force variations (except chatter); so the vibration is less sensitive than force in TCM [26], [212]. In micro machining, vibration is often large and the tiny vibration difference caused by variations of cutting force is relatively negligible [186] [189] [219].

Motor current/Feed Current

The method based on motor current is considered one of the least disruptive and economical methods to estimate feed force and tool state. Li et al [112] [113] [115] presented the approaches to detect the tool wear rate with current. It is accomplished by measuring both feed and motor current and then using a regression analysis that is used to classify the different tool states.

Temperature

Most of the temperature based methods involve some form of thermal imaging system that monitors the chip temperature as 90% of the heat from the engagement point goes into the chip [24][174]. For measurement of the tool tip temperature several methods have been developed as presented by Derrico [40]. It can be used to predict the temperature distribution of the tool face if a proper heat conduction model is used. However, this technique is hard to be implemented in a real cutting process, because it

does not take account the changes in the boundary conditions such as tool geometry and the use of coolant [117].

Vision/Optical

This method can be used in two different ways, either with a camera monitoring the tool tip or the work surface. These two methods can be used as reported by Kurada [98]. He used both the image of the surface texture and that of the tool tip to determine a tool wear state. These methods are shown to work satisfactory but further work has to be done within the area of grey level imaging processing. Oguamanam et al. [147] developed a system extracting five image-based features, which were used in classifying the tool as good, worn or broken in the light of single-point cutting tools.

The sensitivities of the above-mentioned sensors to tool wear are briefly listed in Table 2.1.

Table 2.1 Sensitivity between sensors and tool wear

Sensor	Force	AE	Vibration	Current	Temperature	Optic
Sensitivity	High	High	Medium	Medium	Medium	Medium

2.4 Feature Extraction Techniques

Sensor signals are typically noisy and have to be processed to yield useful features that are highly correlated to tool conditions. This process is called feature extraction. Early work on feature extraction in tool wear monitoring focused on time series methods and frequency domain analyses. An autoregressive (AR) time series model developed by Liang et al. [118] used acoustic emission signal for classification of cutting tool conditions. Autoregressive parameters in a predefined function form are adaptively modified with a stochastic gradient algorithm to provide correlation information. Autoregressive models are also widely used in tool condition monitoring, such as those by Altintas [9], and Kumar et al [97]. Elbestawi et al. [51] employed fast Fourier transform (FFT) to compute the ratio of the magnitudes of cutting force harmonics that are sensitive to flank wear as an indicator of the amount of flank wear. Tarng [191]

conducted spectrum analysis of the displacement signal of the spindle in milling, and found it to be suitable for the recognition of tool fracture. With these methods, a threshold value needs to be set between the normal and abnormal tool states. However, the threshold value varies with cutting conditions. To improve the performance of tool failure sensing, more advanced methods, such as pattern recognition analysis [105] and statistical methods [97], have been developed. These methods have gained various degrees of success in practical applications for monitoring cutting processes. Lee [102] analyzed the relationship of the dynamic force signals obtained in real time by an FFT card with the tool failure of a coated grooved tool, and found that statistical analysis using a t-distribution test showed that the percentage increase of the dynamic tangential force could give a promising threshold for the prediction of tool failure.

The above-mentioned approaches generally assume that the sensor signals are stationary. However, due to the nature of manufacturing processes, the signals are usually nonstationary. The non-stationary nature of signals is due to nonlinearity and/or time dependency of the process [47] [48]. Thus, the approaches that deal with non-stationary signals are more appropriate for process monitoring. Wavelet transform is a convenient tool for processing time varying signals. It has been used to analyze machining signals during tool wear by Tansel et al. [187] [188], Lee and Tarng [102]. Wu and Du [211] introduced the wavelet package transform signal reconstruction for machining process monitoring. Test results both in turning and drilling showed that the proposed method is very effective. Kamrathi at el [87] compared fast Fourier transform (FFT) and fast wavelet transform (FWT) using data from turning force and vibration signals in turning processes. The flank wear estimates were computed from the features extracted through each representation scheme by using recurrent neural network architecture. They showed that the wavelet transform could extract modes of severe tool failure with much greater sensitivity than by spectral analysis. These feature extraction methods are listed in table 2.2.

Table 2.2 Features extraction scopes

	Features	Formula	Reference
Time domain	Mean, Variance Moments, and Cumulants	$\mu = E[x(t)],\ \sigma^2 = E[(x(t)-\mu)^2]$ $Skew = E[(x(t)-\mu)^3/\sigma^3$ $Kurtosis = E[(x(t)-\mu)^4/\sigma^4 - 3$	[144] [105] [143]
	Dynamic/static/ average Component	$\Delta F_a(i) = F_a(i) - F_{med}(i)$ $F_a(i) = \sum_{j=1}^{M} F(i,j) \Big/ M$	[41] [42] [90]
	Normalized Force / Force ratio	$F_i = (F_{a,i} - F_{a,i-1}) \Big/ F_{a,i}$ $F_i = \max[F(i,j)] - \min[F(i,j)]$	[190] [218] [189] [102] [103] [104]
Time series	Coefficients of AR models	$x_t - \sum_{i=1}^{n} \varphi_i x_{t-i} = a_t - \sum_{j=1}^{m} \theta_j x_{t-j}$	[102] [97]
	Residual of AR models	$a(i) = F_{a,i} - \hat{F}_a(i/i-1)$	[9] [8] [58] [47] [188]
Frequency domain	Power Spectrum Density and Harmonics	$S_{xx}(w) = \int_{-\infty}^{+\infty} r_{xx} \exp(-wt)dt$	[159] [51] [41] [87]
Time-frequency domain	Wavelet Coefficients	$c_{j,k} = 2^{j/2} \int_{-\infty}^{\infty} f(t)\overline{\psi}(2^j x - k)dt$	[190] [145] [188] [159] [115] [187] [87] [88]
	Wavelet Package Coefficients	$W_{m,b,n}(i) = 2^{-m/2} \times W_b(2^{-m} i - n)$	[212] [211] [144] [116] [143]

2.5 Tool Wear Classification Methods

Intelligent approaches are attractive when little a priori information is obtained about the system. There is a wide range of methods, including neural networks, fuzzy logic, support vector machines, self-organization maps, and hidden Markov models. The following briefly review the classification methods that have been used for tool condition monitoring.

The estimation of tool wear state from machine signal features is generally achieved with a classifier. Earlier study on TCM is mainly carried out with time series analysis. An autoregressive (AR) time series model developed by Liang et al. [118] used acoustic emission signal for classification of cutting tool conditions. Autoregressive parameters in a predefined function form are adaptively modified with a stochastic

gradient algorithm to provide correlation information. Autoregressive models have also been used in tool condition monitoring, such as those by Altintas [58] and Kumar [97]. With these methods, a threshold value needs to be set between the normal and abnormal tool states. However, the threshold value varies with cutting conditions and is difficult to determine.

To improve the performance of tool failure sensing, more advanced methods have been developed. Neural networks (NNs) are most studied and gained most success in practical applications for monitoring cutting processes, due to its capability in learning and non-linear mapping of features and tool state [150]. Lee [103] developed a MLP to predict flank wear during turning. A neural network scheme is applied to perform one-step-ahead prediction of flank wear by cutting force signals. The input and output were the cutting force ratio and flank wear. Machining experiments conducted using the method indicated that using an appropriate force ratio the flank wear can be predicted to within 7% of the actual wear for various turning conditions. Tansel [186] introduced a Neural-Network-based Periodic Tool Inspector (N2PTI) to evaluate tool condition during the machining of non-metal workpieces. The neural network estimates the tool life from the variation of the feed- and thrust-direction cutting forces. Two neural networks, backpropagation (BP) and probabilistic neural network (PNN) are compared. The advantages of N2PTI are simplicity, low cost, reliability and simple computational requirements.

In case there are only signals but no wear measurement so that there are no training pairs for MLP, unsupervised classification and clustering methods are developed to meet the requirement. Leem [105] designed a sequential Self Organizing Map (SOM) for tool wear monitoring using AE and cutting force signal. This unsupervised neural network achieves high accuracy rates with robustness in the classification of two and three levels of tool wear. Niu [144] applied an adaptive resonance network (ART2) with feature vectors derived from wavelet packet coefficients to identify tool fracture, chipping and chip breakage. At the same time, spectral and statistical analysis techniques have been employed to extract secondary features, which were used as inputs to an ART2 neural network to identify fresh and worn state of the tool. Silva [176] applied two types of neural network, the self-organising map (SOM) and

adaptive resonance theory (ART) to classify the statistical and frequency domain features of the sensor signals. Besides NNs, other pattern recognition methods such as fuzzy clustering approaches [95] [188], linear discriminant analysis [204], Gaussian mixture models [34], combination of regression and neuro-fuzzy techniques [189], and support vector machine (SVM) [220] are also studied and applied to TCM by many researchers.

As stated above, NNs have the advantages of superior learning, noise suppression, and parallel computation abilities. However, successful implementation of an ANN-based monitoring system strongly depends on proper selection of the type of network structure and amount of training data, which are not always available, especially for the abnormal tool state.

Hidden Markov models (HMMs) have recently been employed for TCM lately due to their excellent representation of local dynamics of signals and reasoning property in speech recognition [161][108][23]. The application of the hidden Markov model in TCM was first reported by Heck et al [70] with a continuous density, left-right HMM-based approach for tool wear detection and prediction. It was further studied by Errtunc et al [54], and Baruah and Chinnam [20] in drilling monitoring, and Wang et al [204] and Scheffer et al [172] in turning tool monitoring. Errtunc [54] [55] developed the barograph method and multiple modeling methods to predict the wear status of the drill based on classifying the data signals using a hidden Markov model (HMM). Wang [204] presented a hidden Markov models (HMMs) framework for tool wear monitoring in machining processes. Feature vectors are extracted from vibration signals measured during turning. A codebook is designed and used for vector quantization to convert the feature vectors into a symbol sequence for the hidden Markov model. Experimental results show that successful tool state detection rates as high as 97% can be achieved by using this approach. It was lately successfully applied to milling tool wear monitoring by Owsley et al [150], Atlas et al [15], Fish et al [56] Cetin and Ostendorf [29], Antonio et al [14], and Kassim et al [89]. Fish et al [56] developed a multi-level classifier in milling with two outputs: a prediction of the wear level and a gradient measure that is the posterior probability that the tool is worn. The classifier tracks the dynamics of sensor data within a single cutting pass as well as the

evolution of wear from sharp to dull. It achieved high accuracy across changing cutting conditions, even with a limited feature set drawn from a single sensor.

Table 2.3 lists main references and corresponding classification approaches. The table is not totally exhaustive but serves to be representative of known TCM approaches, where most of them involve wavelet applications.

Table 2.3 Fault classification and estimation scopes

	Approaches	References	Comments
Time series	AR, ARMA	[58], [204], [172], [114]	Linear, Physical meaning, good for stationary machining like turning, not good for non-stationary machining like milling; need set threshold for classification
Principal Component Analysis	PCA/KLT	[118], [97], [169], [154]	Linear, Second order statistics based on eigenvector decomposition, good for Gaussian signal.
Discriminant Analysis	LDA	[78], [202], [212], [204], [172], [174], [169] [214]	Supervised linear classifier, using MSE for optimization, better than PCA for classification, need Gaussian assumption of signal
Neural Networks	MLP	[78] [202] [124] [218], [56] [29] [14] [89] [209], [197] [180]	Iterative MSE optimization; sensitive to network structure; nonlinear classification; slow training
	SOM	[30], [15], [19], [114]	Nonlinear and iterative; suitable for low dimension feature space
	ART	[40], [15], [8], [51], [112], [22]	Based on competitive learning,; fast incremental learning ability, good self-adaptive ability,
	SVM	[220]	Maximizing the margin between classes with minimum number of support vectors; metric dependent, nonlinear, good generalization; slow training
	Others	[44], [43]	Sensitive to training parameters, nonlinear classification, robust to outliers
Clustering	k-means	[78], [202]	K-clusters, the nearest mean decides the cluster, good for Gaussian signal with equal covariance
	Fuzzy Methods	[78], [202], [95], [188], [22], [94]	Need initializing clusters and class membership
	Gaussian Mixture Models	[196], [45]	Each state is assumed to be drawn number of underlying Gaussian distributions; soft membership, better than k-means clustering; need estimate components
Stochastic Models	Hidden Markov Models	[18], [185], [59]	Simple structure; Good in generalization; good in nonlinear and nonstationary machining signals; need to train many small models

AR Autoregressive, ARMA Autoregressive Moving Average
PCA: Principal Component Analysis LDA: Linear Discriminant Analysis
MLP Multilayer Perceptron, SOM Classifying Self-Organising Map
ART Adaptive Resonance Theory, SVM Support Vector Machines
GMM Gaussian Mixture Models HMM Hidden Markov Models

2.6 TCM in Micro-machining and Comments

Compared to conventional machining, there has been little work in micro-maching TCM. The preceding section has presented the different types of difficulties encountered in micro-machining TCM compared with conventional machining. While there is some work reported in micro-drilling tool condition monitoring [95] [188], most studies are on micro-milling and majority are those reported by Tansel and his collaborators [190] [185] [186], [196],[189],[18], [19]. Trujillo et al [196] monitored the micro-milling tool wear with vision approach and found that the intensity of the variation of the reflected laser beam from tool surface is sensitive to tool wear. A wavelet based encoding technique combined with an ART classifier was identified to be effective for this purpose. In [190] Tansel et al applied cutting force to monitor the tool breakage, and found that the variations of static feed force were sensitive to tool breakage. The problem with this approach is that the cutting force is highly noisy. Tansel et al then turned to AE [189] for both tool breakage and 3-state tool wear estimations. Further study by Tansel et al [186] attempted to find exact tool wear value from the cutting forces with MLP neural networks. Motivated by this work, Bao et al [19] [18] developed an analytical force model for micro-end-milling to predict cutting forces and estimate the tool wear. The parameters of the mode were estimated with the cutting force coefficients by genetic algorithm (GA). The model and GA-based parameter estimation approach was tested to be effective and time saving for on-line tool wear monitoring. Tansel et al [185] then used the force model developed in this work to develop a genetic tool monitor (GTM) tool for micro-milling TCM. The tool wear states were estimated from GTM force coefficients based on the experimental data. It was reported that the estimation error was less than 10%.

For the application of NN approaches, however, the monitoring systems strongly depend on the network structure that is hard to generalize, and the amount of training data not always available. An inherent problem of NN is that the weights are sensitive to new inputs and the entire network has to be retrained if new features are added or the states are augmented. These problems were observed in our preliminary studies.

On the other hand, the noise component in the signal for monitoring micro-machining is usually very high and difficult to separate, and this phenomenon also adds to difficulty for TCM in micro-machining [189][190][219]. A noise-robust TCM system is necessary in micro-machining. To overcome these shortcomings, we applied HMM for TCM in this study. Previous application of HMMs for drilling [20] [54] [55] [70] and turning [172] [204] are not suitable to our problems. For HMMs in milling, the important issues for modeling HMMs, such as noise removal, the definition of tool wear states, the selection of the number of HMM states and Gaussian mixture components, were not discussed in [14] [15][29][56][89][150]. The HMM implemented in this study is based on the concept of multi-rate modeling of [56], but with several modifications to adapt to the conditions in micro-milling monitoring and to achieve a noise-robust approach, which is discussed in the later sections accordingly.

Chapter 3
Wavelet Analysis of Sensor Signals with Applications to Tool Condition Monitoring

Effective signal processing techniques are prerequisite for TCM. Wavelet analysis is most suitable for analyzing nonstationary signal processing such as milling forces and vibration. Fourier transform is the basis of modern signal processing and has been the main signal processing methods. It associates the signal from the time domain to the frequency domain and provides useful information. The signal can also be reconstructed by the Fourier transform pair. Wavelet analysis is inherited from Fourier analysis. Many of the important wavelet properties and the design of wavelets are derived from their Fourier transform. In this chapter, we discuss the limitation of Fourier methods, and then introduce wavelet analysis and discuss in depth of various wavelet approaches with applications in TCM.

3.1 Limitation of Fourier Methods: Time-Frequency Resolutions of STFT

We call any square integrable real function $f(t) \in L^2(R)$ a *signal*. For the signal $f(t)$, the Fourier Transform $\hat{f}(\omega)$ is obtained by the inner product of $f(t)$ with a sinusoidal wave $e^{j\omega t}$,

$$\hat{f}(\omega) = \left\langle f(t), e^{j\omega t} \right\rangle = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \qquad (3.1)$$

It transforms the signal $f(t)$ from the time domain to the frequency domain ω and is viewed as the basis of modern signal processing. The Fourier Transform forms the basis of modern signal processing.

The Fast Fourier Transform (FFT) [148] is the standard method for observing signals in the frequency domain. Elbestawi et al. [51] employed FFT to compute the ratio of cutting force harmonics as the indicator of the amount of flank wear. Tarng et al [102] conducted FFT spectrum analysis of the displacement signal in milling, and found it to be suitable for tool fracture recognition. Zhu et al [220] studied the FFT harmonic power distribution of the cutting force signal and developed a thresholding method for fault detection.

In spite of its earlier popularity, Fourier transform has certain serious theoretical drawbacks in processing machining signals. This is because the support of $e^{j\omega t}$ covers the entire real dimension so that $\hat{f}(\omega)$ depends on the value of $f(t)$ for all times $t \in R$. This global inclusivity of information makes it difficult to analyze any local property of $f(t)$ from $\hat{f}(\omega)$. As can be seen from figure 3.1, it has infinite time resolution (upper figure 3.1 (a)), and have infinite frequency resolution (figure 3.1(b)), but it does not has the information regarding the time-frequency relationship. To overcome the limitation of the lack of localization, Gabor [60] introduced a sliding window function $g(t)$ to the Fourier transform and obtains a localized time-frequency atom ϕ:

$$\phi_{u,\xi}(t) = e^{j\xi t} g(t-u) \tag{3.2}$$

The resultant transform is named short-time Fourier Transform (STFT).

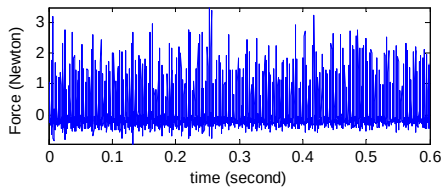
$$(STFT)(f(t)) = \left\langle f(t), \phi_{u,\xi} \right\rangle = \int_{-\infty}^{\infty} f(t) g(t-u) e^{-j\xi t} dt \tag{3.3}$$

The corresponding energy density $\left|STFT(f(t))\right|^2$ is called a spectrogram, which is widely used for time-frequency analysis before wavelet, and applied to TCM [149] [46].

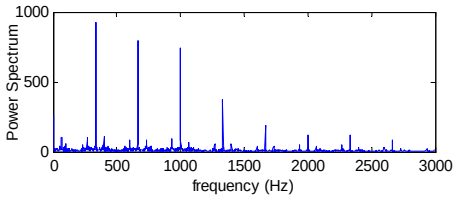
Figure 3.1(c) is the spectrogram of the force from figure 3.1(a). The power spectrum is localized in the time axis and the variation of spectrum energy can be detected in the spectrogram. Once the window is chosen, the frequency resolution σ_ω^2 and time resolution σ_t^2 are constant for both time and frequency, and according to Heisenberg Uncertainty Principle [132],

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4} \tag{3.4}$$

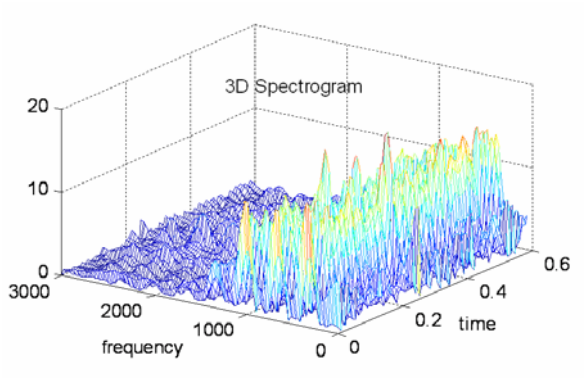
Equality holds if $\phi(t)$ is a Gaussian.



(a): time signal: infinite time resolution



(b) FFT transform: infinite frequency resolution



(c): after STFT transform: fixed time- frequency resolution

Figure 3.1 Cutting forces, Fourier transform, and Spectrogram

It states that Fourier methods can not reach high resolution both in frequency and time. On the other hand, the above mentioned approaches assume that the sensor signals are stationary. However, due to the nature of manufacturing processes, the signals are usually non-stationary [220], [48]. Constant time and frequency resolutions of STFT are not suitable for the analysis of non-stationary signal. For example, Mori *et al.* [139] took the FFT of the thrust force signal and found that the spectra cannot capture the time-localized aspect of the saw-tooth signal, but instead it spread the information across the transformed signal.

The signal processing approaches that deal with non-stationary signals are more appropriate for process monitoring [187]. Gong et al [66] have shown that the wavelet analysis is more sensitive and reliable than the Fourier analysis for recognizing the tool wear states in turning. Yoon et al [214] also verified the reliability of the wavelet transform method compared the spectra method of FFT. Wavelet analysis overcomes the drawbacks of Fourier methods and permits adaptive time-frequency representation.

3.2. Wavelet Analysis

3.2.1 Continuous Wavelet Analysis (CWT)

In wavelet transform (WT), a signal of interest is decomposed into a set of basic localized waveforms, called wavelets. The signal is analyzed by examining the coefficients of the wavelets. To meet the needs for adaptive time-frequency analysis in applied mathematics, physics and engineering, the wavelet theory was developed in late 1980s by Morlet, Meyer, Mallat [131][132], and Daubechies [37], [38]. Wavelet transform is a convenient tool for processing nonstationary signals. It has been widely used to analyze machining signals for tool wear monitoring since Tansel et al [186] [187].

Let $\psi_{a,b}(t)$, $a,b \in R, a \neq 0$, be a family of functions defined as translations and re-scales of a single function $\psi_{a,b}(t) \in L^2(R)$ [132],

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \tag{3.5}$$

with the following basic properties:

$$\int_R \psi_{a,b}(t)dt = 0 \tag{3.6}$$

$$\int_R \psi_{a,b}^2(t)dt = 1 \tag{3.7}$$

These properties indicate that the wavelet is a small wave: oscillating around zero (zero mean) and diminishing quickly (finite energy). Figure 3.2 shows samples of Haar, Morlet and Daubechies wavelets.

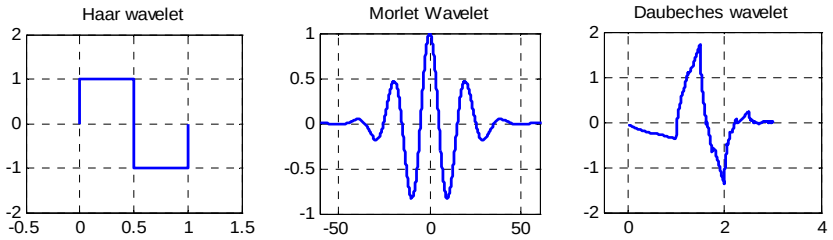


Figure 3.2 Haar wavelet, Morlet wavelet and Daubechies wavelet

Wavelet transform of signal $f(t) \in L^2(R)$ is defined as

$$W_\psi f(a,b) = \langle f(t), \psi_{a,b}(t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi \left(\frac{t-b}{a} \right) dt \qquad (3.8)$$

In physical interpretation, the coefficients of the wavelet transform indicate the variation of energy of the signal with time and frequency. In engineering applications, the square of the coefficients of the CWT is often called as scalogram, defined as Eq. (11), which has been widely used for machinery fault diagnostics, and TCM.

$$SC_t(a,b;\psi) = \left| W_\psi f(a,b) \right|^2 \qquad (3.9)$$

3.2.2 Comparison of Time-Frequency Resolution of FT and WT

The time-frequency resolution of Fourier methods and wavelet analysis are illustrated in figure 3.3. Figure 3.3 (a) is T-F resolution the time signal. As can be seen from the figure, it has infinite time resolution but has no frequency resolution. Figure 3.3 (b) is the T-F resolution for FFT transformed signal. On the contrary, it has no time resolution but infinite frequency resolution. The figure 3.3 (c) is the T-F resolution of STFT. Once the window is chosen, the resolution is fixed. Figure 3.3(d) is the T-F for wavelet analysis. The T-F boxes are adaptive. The 3.3(a)-(c) can be regarded as special case of figure 3.3(d). The signal is localized in the area with time width Δ_t: $[u_0 - \frac{1}{2} a_0 \sigma_t, u_0 + \frac{1}{2} a_0 \sigma_t]$ and frequency width Δ_ω: $[\frac{\eta}{a_0} - \frac{\sigma_\omega}{2a_0}, \frac{\eta}{a_0} + \frac{\sigma_\omega}{2a_0}]$.

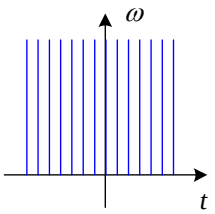
The area of Heisenberg box remains to

$$\Delta_t \Delta_\omega = (a_0 \sigma_t) \times (\sigma_\omega / a_0) = \sigma_t \sigma_\omega ,$$

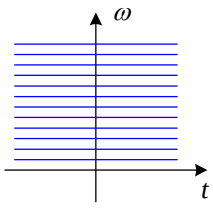
Small a: CWT resolve events closely spaced in time.

Large a: CWT resolve events closely spaced in frequency.

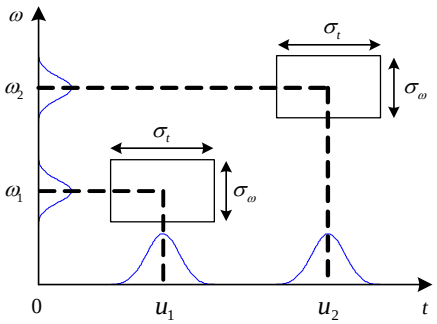
Compared with the STFT, whose time–frequency resolution is constant, the time–frequency resolution of the wavelet transform depends on the frequency of the signal. At high frequencies, the wavelet reaches at a high time resolution but at a low frequency resolution; whereas at low frequencies, high-frequency resolution and low time resolution can be obtained. Such adaptability of time–frequency analysis reinforces the important status of the wavelet transform in the fault diagnostics field.



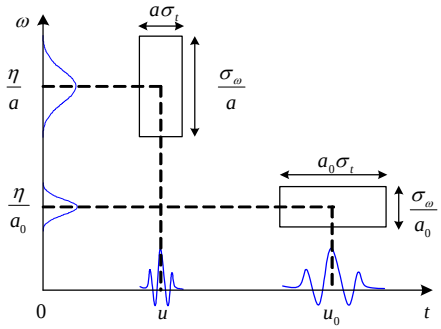
(a): time signal: infinite time resolution



(b): after FFT : infinite frequency resolution



(c): STFT: fixed time- frequency resolution



(d) WT: adaptive time- frequency resolution

Figure 3.3 Time-frequency resolution of and Fourier transform and wavelet transform

3.2.3 Discrete Wavelet Transform(DWT)

Continuous wavelet transforms (CWT) are recognized as effective tools for both stationary and non-stationary signals. However, it involves much redundant information and is computationally very slow. Fast Wavelet Transform (FWT) was developed by Mallat with discrete wavelet transform (DWT) based on Multi-Resolution Analysis (MRA) and constructed with *Conjugate Quadratic Filters* (CQF) [131] [see appendix B for details].

Discretizing $(a,b)=(2^j, k*2^j)$, we have the wavelet function

$$\psi_{j,k}(t)=2^{-j/2}\psi(2^{-j}t-k), \tag{3.10}$$

and define an accompanying scaling function as

$$\phi_{j,k}(t)=2^{-j/2}\phi(2^{-j}t-k) \tag{3.11}$$

at level of detail *j* and shift *k*.

The signal $f(t)$ is represented by shifted and dilated versions of $\psi(t)$ and $\phi(t)$ with DWT,

$$f(t) = \sum_{k \in Z} c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_{k \in Z} d_{J,k} \psi_{J,k}(t) \qquad (3.12)$$

where

$$c_{j,k} = 2^{-j/2} \int_R f(t) \phi(2^{-j} t - k) dt \qquad (3.13)$$

is the scaling coefficient, which measures the local mean around the time $2^{-j} i$, and

$$d_{j,k} = 2^{-j/2} \int_R f(t) \psi(2^{-j} t - k) dt \qquad (3.14)$$

is the wavelet coefficient, which characterizes the local variation around the time $2^{-j} i$ and the frequency $2^{-j} f_0$.

Figure 3.4(a) illustrates the 5-level MRA analysis of a cutting force signal sampled at 6000Hz. With a 5-level MRA analysis, the corresponding frequency bands are separated as illustrated in figure 3.4(b). By DWT, the signal $f(t)$ is decomposes into two parts: low-pass approximation coefficients and high-pass detail coefficients. The next step then decomposes the new approximation coefficients.

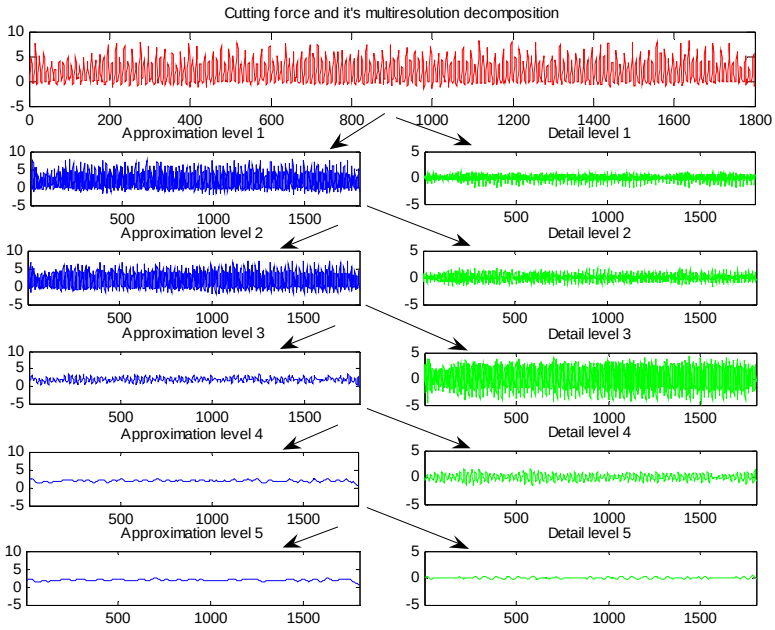


Figure 3.4(a) 5-level MRA analysis of cutting force

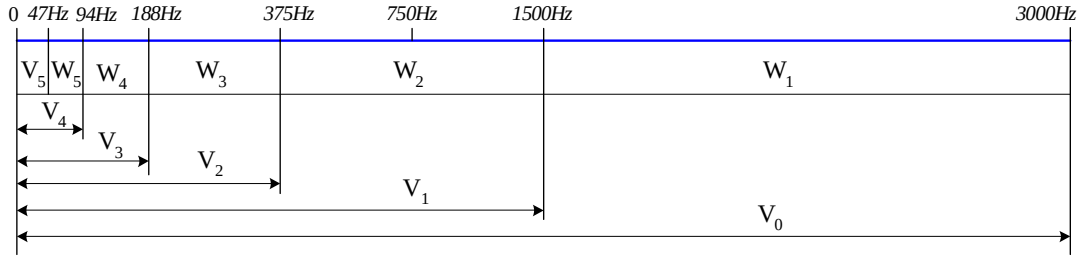


Figure 3.4(b) Frequency band separation of 5-level MRA analysis

3.2.4 Wavelet Packets Decomposition

The DWT may lead to a loss of useful information at high frequency, since successive details are no longer analyzed. We need to double the sampling rate for higher frequency analysis, which however involves more data and computation. Wavelet packet decomposition [208] is a generalization of wavelet decomposition at higher frequencies. In the wavelet packet decomposition, each detail coefficient vector is also decomposed into two parts using the same approach as DWT in approximate coefficients. This offers the analysis at the high frequencies.

Basically, in WPD, for fixed values of j and k, $W_{j,n,k}$ analyzes the fluctuations of the signal roughly around the position $2^j \cdot k$, at scale 2^j and oscillation parameter n.

WPA operates by approximating a signal with scaled and translated wavelet packet functions.

$$W_{j,n,k} = 2^{-j/2} W_n(2^{-j/2} t - k) \qquad (3.15)$$

The notation which characterizes each wavelet packet $W_{j,n,k}$ reflects the scale 2^j and location $2^j k$.

The wavelet packet coefficients are then produced from the integral:

$$d_j^k = \int W_{j,n,k}(t)\, f(t)\, dt \qquad (3.16)$$

It should be emphasized that equation (3.15) allows many possible combinations of wavelet packet functions to be selected in order to optimally characterize the signal.

Several criteria have been developed to select the best basis for these purposes [132] [208].

Due to the benefits of wavelet decomposition, wavelet methods have been studied in all aspects in TCM, such as time frequency analysis, signal denoising, feature extraction and compression, or directly used as classifier for TCM. Basic theories of these approaches are introduced and literatures are reviewed in the following sections.

3.3 Applications

The most important property of wavelet useful in tool condition monitoring is its sparse representation of signals. The wavelet expansion coefficients $c_{j,k}$ and $d_{j,k}$ decay rapidly with increase in j and k, and only a few large coefficients exist while the others are small. By setting a suitable threshold, we can filter the undesired noise. This is the essence of wavelet denoising, compression and detection. Donoho et al [43], [44] showed that denoising by soft thresholding is nearly optimal. Another important property of wavelet transform that determines its applications is its localization property, as discussed in section 3.2. Unlike Fourier transform that spans the entire time period, wavelet transform localize the time and frequency description of the signal, and reveals the signal behaviour in certain time and its corresponding frequency property, which is generally useful for uncovering for different localized features associated with various different tool states.

We first presents a review of the time-frequency analysis of machining signals, signal denoising and feature extraction of wavelet applications for TCM. These three approaches are studied in most of the literature. Two new approaches, called singular detection and density estimation, with wavelet for TCM are also introduced with real experiments and discussed for TCM. Though few papers are found on these two approaches on TCM, papers related to machinery condition monitoring and fault diagnosis are reviewed to show the prospect of these applications in TCM. Table 3.1 lists the references reviewed according to different wavelet features used for TCM in this chapter.

Table 3.1 Overview of the selected references with wavelet applications

Wavelet	Wavelet Features	Tool Condition	Monitored Signal	Reference
CWT	Scalegram: $SC_t(a,b;\psi)=\left\|\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt\right\|^2$ Mean, Variance of Scalegram.	Drill breakage	Current	[180]
		Drill wear	Current	[180]
		Mill breakage	vibration	[52]
		Milling chatter	vibration	[41]
DWT	$c_{j,k}=2^{-j/2}\int_R f(t)\phi(2^{-j}t-k)dt$ $d_{j,k}=2^{-j/2}\int_R f(t)\psi(2^{-j}t-k)dt$ Mean, Variance Moments, and Cumulants of DWT	Drill breakage	Current	[180],
			Force	[173], [8], [51], [39],
			AE	[22], [94]
		Drill wear	Force	[195], [7]
			Vibration	[179]
			AE	[94]
			Curremt	[180]
		Milling chatter	Force	[117], [125], [6]
		Milling tool wear	Force	[150], [90]
		Milling fracture	Current	[214]
		Turning tool wear	AE	[43] [44],
			Force	[[24],[213], [196]
			Vibration	[208], [59]
		Turning failure	Force	[66]
		Grinding too wear	Force	[139]
WPD	$d_j^k=\int W_{j,n,k}(t)\,f(t)\,dt$ Mean, Variance Moments, and Cumulants of WPD	Drilling wear	Vibration	[192]
		Drilling breakage	Current	[187]
		Turning chatter	Force	[192]
		Turning chipping/ breakage	AE	[40], [115]
		Turning wear	AE	[40], [115]
			Current	[153]
			Vibration	[114]
			Force	[114],[22]
MP	$f=\sum_{n=0}^{m-1}\left\langle R^n,\phi_{\gamma_n}\right\rangle\phi_{\gamma_n}+R^m$	Drilling failure	Force	[102]

3.3.1 Time-Frequency Analysis of TCM Sensor Signals with WT

In the time-frequency analysis of TCM sensor signals, the scalogram $SC_t(a,b;\psi)$, wavelet coefficient $d_{j,k}$, scaling coefficient $c_{j,k}$, wavelet packet coefficients d_j^k, and

their wavelet domain statistics (i.e: mean, variance etc.) are used as criterion in condition discrimination in TCM after some manipulation.

Yesilyurt [213] presented the use of the scalogram and mean frequency of scalogram of vibration signals in milling. The only variable is the feedrate while other cutting parameters are fixed. The mean frequency of scalogram characterized the energy density of the signal in a certain period. It was found that the feed rate is highly correlated to the mean frequency of scalogram, and the mean frequency variation is quite sensitive to the presence of tool wear.

With CWT, we generally encounter the problem of overlapping and may smear the localized features for TCM. Minimizing the effect of overlapping and improving the localization are still problems in CWT. These are partially studied in [196] with an exact WT for gear fault detection. DWT is preferable in this aspect. Gong et al [66] applied DWT to monitor the flank wear states in turning. It was found that the 5-level coefficients were sensitive to the flank wear and cutting conditions. Fu *et al.* [80] identified sawtooth and screeching behavior in the thrust force signal by using convolution masks to extract various features from the DWT coefficients. The drawback of this method is that the profiles of the convolution masks are derived from idealized simulated signals, and the shapes of the simulated signals significantly influence the output results. Suh [179] developed a DWT based algorithm to capture the variations of periodicity, scaling coefficients, and wavelet coefficients. The three parameters were used to characterize stable, conditionally stable, and unstable milling conditions. It was claimed that this approach can accurately detect the transition of the system from one state to the other. Yoon et al [214] applied with the standard deviation ratio of DWT coefficients of cutting force to detect chatter. It was found that the detail coefficient parameters of the third or fourth level were desirable for detection of chatter with spindle speed of 500-1,300rpm in end milling. It was reported in [150] that the variance of wavelet coefficients are sensitive to tool wear and little influenced by the variation of working conditions, and thus provide a robust description of tool wear in milling.

Wavelet transform can be very helpful if it is used as a signal separation tool. Liu [122] presented a method to find different signal components with WPD. This method

divides the signal vector space into two subspaces: One subspace contains the transient components excited possibly by localized defects, and the other contains the remaining components. This basis is adapted to reveal the characteristics of the transients and other harmonic components. It overcomes the limitation of the Coifman and Wickerhauser's best basis algorithm [208] in the detection of transients and is more effective in machinery condition monitoring, especially in the early stage of fault development. Niu et al [143][144] separate the acoustic emission signal into burst and continuous components by WPD. The burst signal is suitable for the detection of transient tool conditions, such as chipping and tool fracture, etc., and the continuous signal is suitable for determination of tool wear. This is promising as only one signal is used to carry different monitoring tasks.

The WT can detect localized time-frequency of machining signals, but it generally needs a feature extractor followed by a classifier for TCM. The Matching Pursuit (MP) [134] algorithm handles both of these problems in a straightforward manner. It decomposes the signal into a linear expansion of waveforms that are selected from a dictionary of basis functions. It chooses waveforms that best match the signal structure iteratively. As a pattern recognition method, MP adapts to specific machining monitoring tasks and do not need feature extraction or classification. Fu et al [80] applied the MP to predict small drill bit failure, with different types of drilling behavior from the thrust force. It was revealed that the MP approach performed satisfactorily with Gaussian, Haar, and Gabor wavelet dictionaries in detecting small drill bit pre-failure. This approach simplifies the need of a classifier for TCM and is promising for future studies. The problem is that the threshold selection in the wear detector is hard to generalize.

In some cases, the energy variation of non-stationary signal for TCM is examined. Wavelet (also STFT) can describe local properties of nonstationary signal but the scalegram and spectrogram can only represent the energy distribution roughly. Quadratic time-frequency methods such as Wigner-Ville Distribution (WVD) and Choi-Williams Distribution (CWD) [46] are more appropriate. Bradford and Atlas [105] stated that WVD provides good stationary and non-stationary representation, and provide alternatives for time-frequency analysis in TCM. These methods are widely

applied to mechanical signal analysis for machinery condition monitoring [177], [46], and shows good time-frequency energy distribution for fault detection.

Table 3.2 summarizes the time-frequency properties approaches reviewed in this section.

Table 3.2 time-frequency properties of the transforms

Wavelet	Reference	Transform	Suitability	Comments
CWT	[73-75]	linear	Non-stationary signal	Constrained to UP, can not reach both high time and frequency resolution; redundant transform, computationally slow
DWT	[62], [76-81]	linear	Non-stationary signal	Constrained to UP, can not reach both high time and frequency resolution; only low frequency components iteratively decomposed
WPD	[24], [60], [82], [83], [121], [127], [130]	linear	Non-stationary signal	Constrained to UP, can not reach both high time and frequency resolution; redundant transform; both low and high frequency components iteratively decomposed
MP	[85]	linear	Non-stationary signal	Constrained to UP, can not reach both high time and frequency resolution; acted also as a classifier
STFT	[49], [58], [62], [63]	linear	Stationary signal	Constrained to UP, can not reach both high time and frequency resolution
WVD/ CWD	[86-89]	quadratic	Non-stationary signal	Not constrained to UP, accurate time-frequency distribution, but involves cross-interference terms

UP: Uncertainty Principle

3.3.2 Wavelet Thresholding for Denoising

Signal noise always exists in machining. Denoising is a common practical problem in signal processing. Ali et al [5] found that the spectra of the AE signal are highly contaminated by noise. This noise is so large that the first principal component had no discriminatory power at all. But denoising is generally not concerned in TCM. This may due to the high signal to noise ratio (SNR) in conventional machining, and the noise imposes little effect on the final decision. And in statistical analysis such as mean or moving average [74], the noise is diminished by the averaging. But in high precision and micro machining, the machining signal is typically very small, and as a result the SNR is relatively low [218]. The noise has to be removed before further analysis for TCM.

The model of signal with noise is [81],

$$y(t_i) = f(t_i) + \varepsilon(t_i) \tag{3.17}$$

where the function $f(t)$ represents the desired signal, while the remaining part ε_i is independent and identically distributed (i.i.d.) noise. The noise-free signal $f(t)$ is unknown and we look for an estimator $\hat{f}(t)$ that will remove the noise.

For the wavelet denoising, apply DWT to the noisy data, the wavelet coefficients:

$$d = Wy(t_i) = Wf(t_i) + W\varepsilon(t_i) \tag{3.18}$$

Because smaller coefficients are usually contributed by data noise, thresholding out these coefficients has the effect of removing the data noise. In wavelet thresholding, after setting some coefficients to zeros, the reconstructed (denoised) signal is obtained by inverse transformation.

$$\hat{f}(t) = W^{-1}\hat{d} \tag{3.19}$$

Figure 3.5 illustrates the Thresholding and reconstruction process.

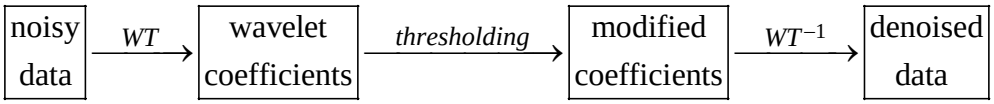


Figure 3.5 wavelet denoising scheme

Donoho and Johnstone developed several wavelet-based thresholding techniques such as hard thresholding and soft thresholding [43] [44] to find an optimal estimate $\hat{f}(t)$ from the noisy data. The hard thresholding employs a "keep-or-kill" rule (figure 3.6a), while the soft thresholding is a "shrink-or-kill" rule (figure 3.6b). To remedy the drawbacks of both hard (unstable, and sensitive to small changes) and soft thresholding (bias estimate) rules, Gao and Bruce [61] developed the *firm* threshold thresholding (figure 3.6c). The resulting wavelet thresholding offers a balance between the two approaches. Note that firm thresholding has a drawback in that it requires two threshold values, thus making the estimation procedure for the threshold values more computationally expensive.

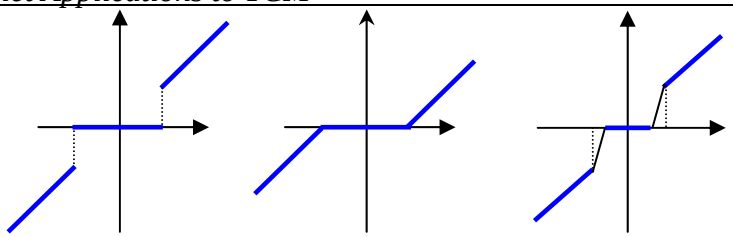


Figure 3.6 a) Hard thresholding b) Soft thresholding c) Firm thresholding

The choice of threshold λ is crucial: small/large threshold values will produce estimates that tend to overfit/underfit the data. Donoho and Johnston [43] proposed a universal threshold $\lambda = \hat{\sigma}\sqrt{2\log(n)}$, where n is the number of observations and $\hat{\sigma}$ is an estimate of the noise variance, which is unknown and needs to be estimated from noisy samples. Despite the triviality of such a threshold, they showed that the resulting wavelet estimator is asymptotically near-minimax among all estimators within the whole range of the Besov space[45]. This threshold is only dependent on the number of data and data variance, but not dependent on the decomposition levels, which is a not practical in many real problems. A way to overcome this problem is to estimate level-dependent thresholds with Generalized Cross-Validation [141], or utilize the a prior information with Bayes shrinkage [1], [2] for signal denoising. (See Jensen [81] for a comprehensive study).

Wavelet denoising is widely employed in machinery condition monitoring and fault diagnosis. Menon et al [138] used the wavelet-based method to eliminate the background noise, which is a problem when using the AE to detect small fatigue cracks in rotor head components. Bukkapatnam et al. [25] modified Donoho's thresholding method for chaotic signal with multiplicative noise. It was found that the method can separate chaotic signal from worn tool in machining, and suitable for on-line implementation. Jeong et al [83] proposed methods that demonstrate their competitiveness to the existing engineering data-compression and statistical data-denoising methods for achieving the data reduction goals. Further experimentation with a tree-based classification procedure for identifying process fault classes illustrates the potential of the data-reduction tools. Qiu et al [159] compared the performance of wavelet decomposition-based de-noising and wavelet filter-based de-noising methods on signals from mechanical defects. The comparison result

demonstrates that wavelet filter is more suitable and reliable to detect a weak signature of mechanical impulse-like defect signal, whereas the wavelet decomposition de-noising method can achieve satisfactory results on smooth signal detection. Lin and Qu [120], and Lin et al [119] improved the soft-thresholding method by utilizing a prior information on the probability density of the impulse, which matched with Morlet wavelet. The time-frequency resolution can be adapted to different signals of interest. It was claimed that the method performed excellently when applied to de-noise gear and bearing vibration signals with a low SNR.

In TCM, Kwak and Ha [99] described the use of the grinding force signal with noise reduction to detect the dressing time based on DWT. As a result of de-noising, the grinding force signal was successfully used to detect the need for dressing. The wavelet de-noising method was found to be more effective than the FFT filtering technique. Kwak et al furthered this approach in turning [100]. DWT is used in both de-noising and detecting tool failure. The DWT coefficients of the cutting force signal showed that the onset time of tool failure and chatter vibration was successfully detected. Li and Guan [111] proposed a wavelet-based de-noising to extract marked features from the feed-motor current signals to indicate the minor cutting edge fracture. It was found that the best de-noising approach was to utilize a third symmlet mother wavelet function in combination with cross-validation threshold determination and soft thresholding.

It has to mention that the existing wavelet de-noising methods reported in the literature rely heavily on white Gaussian noise and relative energy levels of wavelet coefficients. The wavelet denoising methods meet problems in denoising non-Gaussian noise, which is discussed in [12][16]. It was found that the thresholds defined by [43] and [44] are too small when we met Super-Gaussian (normalized kurtosis>0) noise. In this condition, the noise can be regarded as a certain signal and the wavelet decomposition coefficients are not evenly distributed among all scales as Gaussian noise, and as a result some of the noise coefficients are not small. Under this condition, independent component analysis (ICA) [36] [75] for noise separation is more effective than wavelet denoising as shown in [217][218]. Chapter 5 will discuss this problem in details.

3.3.3 Feature Extraction and Dimension Reduction

Apart from the original intention of the wavelet transform for time-frequency analysis of non-stationary signals, the important and successful application of wavelet in TCM is fault feature extraction and dimension reduction. These wavelet coefficients usually need to be extracted for robust and effective representation of different tool states in TCM. The extraction approaches are generally based on some statistical measurements (i.e. mean, variance, Euclidian distance, etc.) to maximize the discriminability or classification among different tool states.

A good example of wavelet feature extraction is presented by Wu and Du [210]. They introduced an automatic feature extraction and assessment procedure using a wavelet packet transform in TCM. To identify the effectiveness of the selected features in both time and frequency domains, four criteria such as cross-correlation and cross-coherence of signal and reconstructed signal, correlation of the residue, and power spectrum of the residue, are proposed. The proposed method is tested for chatter monitoring in turning and tool wear monitoring in drilling. It was found that WPD can capture important features of signal that are sensitive to the machining conditions, e. g., chatter and different states of tool wear, and insensitive to the variation of working conditions and noises. Accordingly, accurate and reliable on-line monitoring decisions can be made. Wu et al [211] furthered the study with a real-time implementation of WPD to motor current signal. The wavelet packet features were obtained under normal tool condition. Based on the principal components analysis (PCA) the alarm thresholds are then constructed for decision making.

Moria et al [139] developed a method for extracting pre-failure features from the cutting force to predict breakage of a small drill bit to discriminate different tool states. The DWT coefficients are reduced to three indices: energy, waviness and irregularity. The problem is that it is difficult to determine the index function for classification. Tansel *et al.* [186] used the DWT coefficients of the thrust force as an input to (ART-2) to predict micro drill bit breakage in peck drilling. The authors later enhanced this technique by encoding the raw data before inputting it to the ART-2 [188]. However, the encoding technique relies excessively on statistical abstractions that cannot be explained in terms of the machining system behavior during failure. Thus, the drilling

monitoring method lacks a basis for generalization to other types of drilling processes. Hong et al [72] decomposed dynamic cutting force signal into different frequency bands by DWT, and two features, mean values and variances of the local maxima, were extracted from the decomposed signal for each frequency band. These features are input to MLP for the tool state classification. It was shown that the features extracted by wavelet transform are less sensitive to changing cutting conditions and the MLP has high classification rate.

Kamarthi et al [87] [88] investigates a flank wear estimation technique in turning through wavelet representation of acoustic emission (AE) signals. The energies of DWT coefficients are used for flank wear estimation. The flank wear estimation from the recurrent neural network (RNN) is fairly accurate and indicates that the FWT representation of AE signals is more effective and sensitive than Fourier transform representation. Pittner and Kamarthi et al [158] extended this study with a new feature extraction method based on the FWT. The features of AE signal are calculated by the Euclidean norms of the frequency clusters. They showed that the proposed method can efficiently extract important features related to progressive flank wear. Choi et al [125] studied cutting force trends and tool wear effects in ramp cut machining. This study is challenging because the depth of cut is continuously changing in ramp cuts. Wavelet analysis is applied to cutting forces from a progressively worn tool. The root mean square (RMS) value of the approximation coefficients extracted as features and linear regression are used for tool wear estimation. It is reported that for small depth of cut, the linear regression can estimate the tool wear with error below 6%. Abu-Mahfouz [4] found that the average harmonic wavelet coefficients and the maximum entropy spectrum peaks are more efficient in training the MLP than the time domain features in monitoring of twist drill wear. Other features, such as normalized mean wavelet packet coefficients of the feed force, which is affected not only by the flank wear but also by the severe crater wear observed in high speed machining, are tried by Obikawa and Shinozuka [145]. As these high-dimension features need to be reduced for less computation and robust representation, an algorithm is implemented with *Fisher's discriminant ratio (FDR)* [192] to discriminate features from different classes.

Table 3.3 summarizes the feature extraction approaches reviewed in this section.

Table 3.3 Feature extraction with wavelet analysis

Features	Reference	Comments
DWT	[61], [67], [68], [120], [122-126], [128], [129], [131]	Statistics of DWT coefficients; need succeeding classifier or set a threshold for classification
WPD	[24], [60], [82], [121] [127], [130], [132-136]	Statistics of WPD coefficients; need succeeding classifier or set a threshold for classification
MP	[85]	Acted as both feature extraction and classification

3.3.4 Singularity Analysis for Tool Wear Detection

Singularities can be thought of as either an abrupt change or 'impulse' in a signal, or the sudden shift of the signal's mean value to a different level. The good time-frequency localization property provides wavelet in singularity analysis [133]. The main goal Singularity analysis is to estimate the number, localization and sizes of a signal's abrupt changes and spikes or image edges [136]. The general idea used to detect a function's abrupt changes through wavelet coefficient regularity at certain point and the rate of decay of the coefficients near this point. It has been studied in machinery condition monitoring lately, but little is studied in TCM. Because it is robust and stable in capturing signal changes, it can be expected that these singularity based features will provide valuable study in TCM.

The singularity of the signal is measured by the Lipschitz Exponent (L.E.). Assume the signal $f(t)$ has the following property at the vicinity of t_0

$$\left| f(t_0 + h) - P_n(h) \right| \le A|h|^{\alpha}, \; n < \alpha < n+1 \tag{3.20}$$

where $P_n(t)$ is a n-order polynomial and pass the point $f(t_0)$, and h is small enough, the Lipschitz exponent (L.E.) of $f(t)$ at t_0 is the superemum of the α for which function (3.20) holds. It follows directly that if α determines $f(t)$'s continuously differentiable times at t_0. The Lipschitz exponent (L.E.) α gives an indication of how regular the function $f(t)$ at t_0. The higher the α, the more regular the function $f(t)$ is.

Mallat and Hwang [133] have shown that the local regularity of a function (signal) is related to the propagation across scale of its wavelet maxima and the decay of the wavelet transform modulus in the time-scale plane. The definition of the maxima along scale:

$$\frac{dWf(s,t)}{dt} = 0 \tag{3.21}$$

The modulus maxima line consists of the points that are local maxima in the time-scale plane. Along the modulus maxima line the wavelet coefficients have the scaling behaviors as follows:

$$\log_2 \left| WT_{2^j} f(t) \right| \le \log_2 A + j\alpha \tag{3.22}$$

This function connects the wavelet scale j and the Holder exponent α. The A and α can be computed by setting the equality in function (3.22).

Beyond the main application is image edge detection, the idea of singularity analysis with wavelet is implemented in [13] [146] for change point analysis, [17] for underwater transient analysis, and [178] for non-linear system identification, which is similar to the TCM in machining. Hambaba et al [67] uses wavelet transform to determine the Holder exponent value of a gear response at different scale levels. By fitting an autoregressive moving average (ARMA) model to the wavelet-transformed data, analysis of the residual error is used to indicate the presence of damage in the gear. Sun et al [182] proposed singularity analysis for bearing defect diagnosis. Through modifying the intensity of the wavelet transform modulus maxima, defect-related vibration signature is highlighted and can be easily associated with the bearing defect characteristic frequencies for diagnosis. Loutridis and Trochidis [121] employed the Lipschitz exponent to investigate gear faults. The Hoelder exponents for each type of faults were calculated from the resulted modulus maxima lines. It was shown that the Hoelder exponent for each type of fault exhibits a constant value, not affected by changing working conditions, and can be used as a stable indication of different gear faults. It has also been showed that the Holder exponent is effective in capturing non-stationary nature of signals and sensitive for identifying structure damage [146]. Peng et al [152] examine shaft orbits using the wavelet modulus maxima. Four L.E. based features are extracted to classify the shaft orbit. Peng et al [154] furthered this study

and present a novel singularity based fault feature extraction from vibration signals. The data are collected under different machine health conditions that show different patterns of singularities measured quantitatively by the Lipschitz exponents. The number of Lipschitz exponents per rotation, the mean value and the relative standard deviation of the Lipschitz exponents that are obtained from the extracted features for singularity representation. The results show that the three parameters are excellent fault features for fault pattern recognition.

Figure 3.7 demonstrates an example of with modulus maxima for detection of milling tool at different wear states. Fig. 3.7a shows a cutting force signal from a fresh tool and severe worn tool with their wavelet modulus maxima. Lipschitz exponent (L.E.) extracted from the modulus maxima line (figure 3.7b). Though they do not change much, Lipschitz exponents offer a stable tool to detect the degrees of tool wear level as shown. More tests and their statistics of LE from two tool states are illustrated in figure 3.7c. It is observed that with the fault severity increasing, the milling force signals' singularities and singularity ranges will increase as well, and therefore we can evaluate the fault severity through measuring the force signals' singularities and singularity ranges.

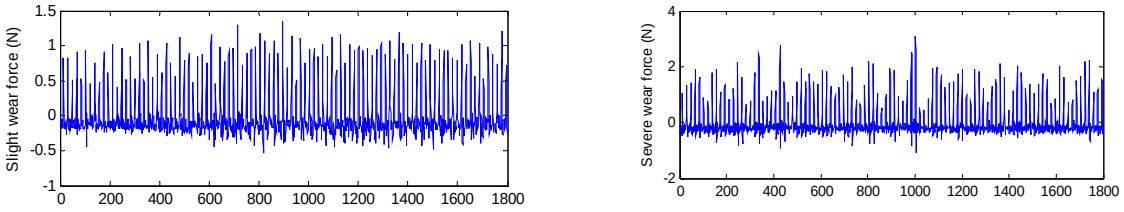


Figure 3.7 (a) cutting forces of slight flank wear and severe flank wear

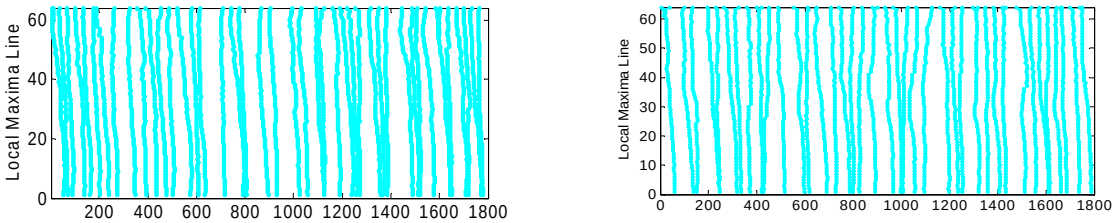


Figure 3.7 (b) Modulus maxima for slight wear tool and severe wear tool

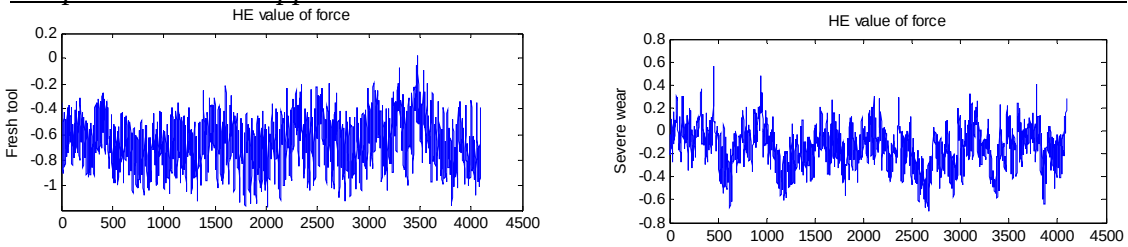


Figure 3.7 (c) Lipschitz exponent of slight wear tool and severe wear tool

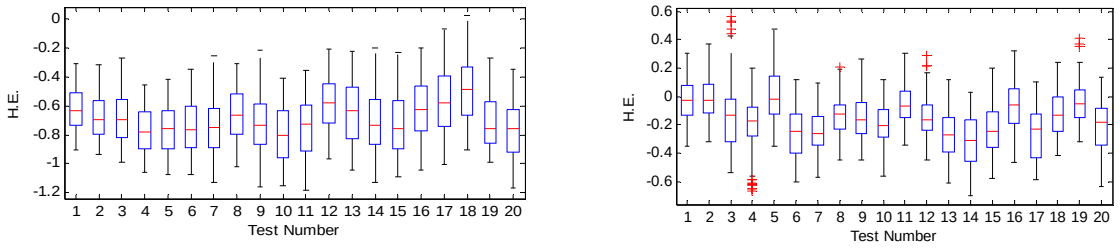


Figure 3.7 (d) Boxplot of more tests of Lipschitz exponent

Figure 3.7 Modulus maxima for slight wear tool and severe wear tool

3.3.5 Wavelet Density Estimation for Tool State Classification

Density estimation is a classic problem in statistics [3] [200]. The density estimation problem in statistics is in fact the pattern recognition problem in engineering: assigning classes to higher probability density groups. The most important and popular density estimation approach before wavelet density estimation is kernel density estimation [69]. Wavelet is brought attention in density estimation lately [199]. The differences between kernel and wavelet estimates are mainly explained by the ability of the wavelet method to take into account local gaps in the data distribution. This new approach is very promising, since smaller structures superimposed onto a larger one are detected by this technique, especially when small samples are investigated. Thus, wavelet solutions appear to be better suited for classification studies.

Wavelet density estimation is developed from non-parametric regression problem in statistics. Wavelet density estimation opens the problem of density estimation for dependent observations, while almost all other density estimators fail in dealing. Let $X_1, X_1, ...X_n$ be a sequence of independently and identically distributed (*iid*) random variables, with a common density function: $h = h(x)$, and it is express in the wavelet series.

$$h(x) = \sum_{k \in Z} a_{j,k} \phi_{j,k}(t) + \sum_{j=-\infty}^{J} \sum_{k \in Z} d_{J,k} \psi_{J,k}(t) \tag{3.23}$$

To estimate h, it is sufficient to estimate the coordinates $a_{j,k}$ and $d_{J,k}$.

Usually this expectation is estimated by the sample mean:

$$\hat{a}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(X_i) \tag{3.24}$$

Similarly,

$$\hat{d}_{J,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{J,k}(X_i) \tag{3.25}$$

Inserting this estimates to the wavelet series expansion of h, we get an estimate:

$$\hat{h}(x) = \sum_{k \in Z} \hat{a}_{j,k} \phi_{j,k}(t) + \sum_{j=-\infty}^{J} \sum_{k \in Z} \hat{d}_{J,k} \psi_{J,k}(t) \tag{3.26}$$

The estimate $\hat{h}(x)$ is then used as real probability density for application.

Muller and Vidakovic [140] used wavelet bases functions to build a series expansion of an unknown density. A prior model imposed on the wavelet coefficients induces geometrically decreasing probabilities on non-zero coefficients at finer levels. Abramovich and Angelini [2] imposed prior on each wavelet coefficient is a mixture of a point mass at zero and a Gaussian density. This approach is more flexible since it can approximate wider densities. Heck and Chou [71], and Chou and Heck [34] reported a study similar to wavelet density estimation for TCM. They use a Gaussian Mixture Model (GMM) to approximate the wavelet coefficients from machining signals. Different tool states are represented with different GMM, and then choose the maximum class conditional probability for state decision when feature extracted for classification. Safavi et al [167] present an application of wavelets analysis to density estimation and process monitoring. The resulting density function was used to define a normal operating region for the process so that any future abnormal changes in the process can be monitored. Results of applying these techniques to chemical process are also presented to be effective in monitoring the states.

There is no study of this application to TCM at present. However, this idea can be easily adapt to tool wear classification with the following steps:

1) Estimate the respective density distributions of tool wear levels provided by the machining data and measured tool wear, like the supervised neural network training;

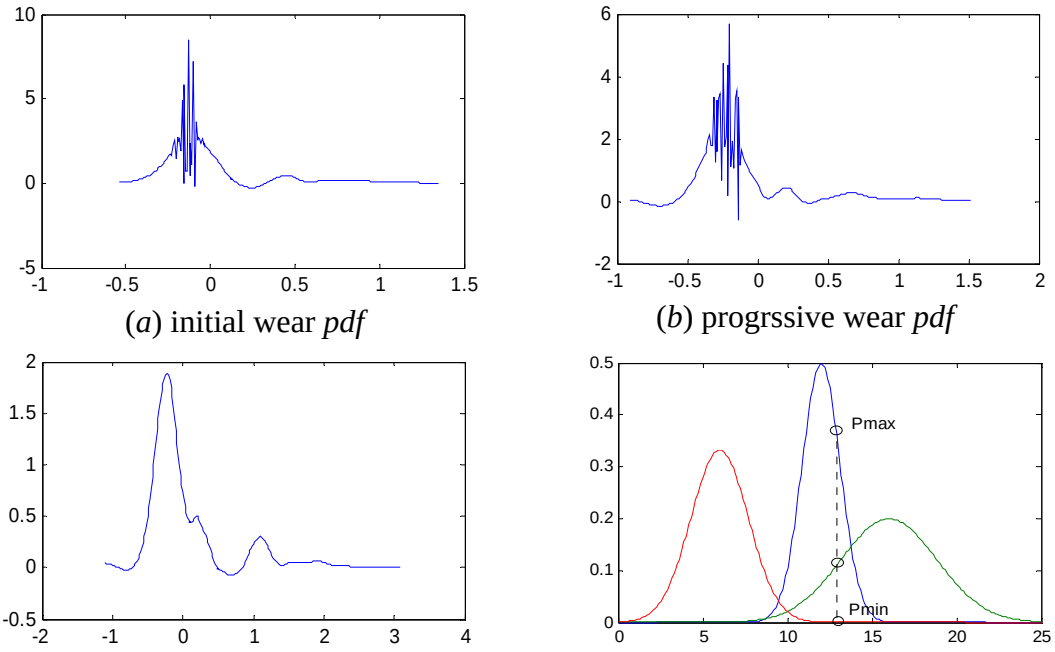
2) Match the present estimated signal density to the known density;

3) Select the maximum one as tool condition with Bayes classifier.

The Bayesian classifier then classifies the state by maximize the posterior probability density (Maximum a Posterior, MAP, in Bayesian literature).

$$W_i^* = \arg\max_{w \in W} \{ p(W_i = w \mid y_1, ... y_i) \} \tag{3.28}$$

Figure 3.8 (a),(b),(c) are the learned density from the measurement, representing three possible tool state. The estimated density is matched to these tree conditions, and the likelihood of these matching is found in figure 3.8(d). We conclude that the state highest probability is what we estimate.



(*a*) initial wear *pdf* (*b*) progrssive wear *pdf*

(*c*) accelerated wear *pdf* (*d*) Bayesian classifier (MAP)

Figure 3.8 Density estimation with wavelet and state estimation

3.4 Conclusion

Wavelet theory and applications of wavelet analysis to TCM are introduced and reviewed in this chapter. Because of nonstationary property of machining process, wavelet is more effective than Fourier methods in tool condition monitoring signal analysis. Based on the sparsity and localization properties of wavelet decomposition discussed, applications of wavelet in TCM are introduced for cutting force time-frequency analysis, denoising and feature extraction. Prospects of wavelet applications in TCM with singularity analysis and density estimation for tool state classification are also discussed and case studies are presented to show their effectiveness.

Chapter 4
Framework of TCM

4.1 TCM as a Pattern Recognition Problem

The problem of TCM can be considered as a typical pattern recognition problem. For pattern recognition, a given pattern is to be assigned to one of C categories $\omega_1, \omega_2, ...\omega_c$ based on a vector of d feature values $y = (y_1, y_2, ..., y_d)$. The features are assumed to have a probability density function conditioned on the pattern class. Thus, a pattern vector x belonging to class ω_i is viewed as an observation drawn randomly from the class-conditional probability function $p(y \mid \omega_i)$. We can formally specify the objectives of TCM to be a search for the most probable state ω_i given the extracted measurable signal feature y. This is a dynamic inference problem since we do not estimate the tool state only with prior knowledge, but also adapt to the current features. This is somewhat of Bayes inference [49]: assign input pattern y to class ω_i if

$$P(\omega_i \mid y) > P(\omega_j \mid y) \quad for\ all\ j \neq i \tag{4.1}$$

Hence, the aim of TCM is to find,

$$\text{TCM: } \arg\max_i p(w_i \mid y) \tag{4.2(a)}$$

or in the physical form

$$\text{TCM: } \arg\max_{tool\ state\ i} p(tool\ state \mid signal\ features) \tag{4.2 (b))}$$

A number of state estimation approaches such as Bayes decision rule, neural networks, clustering, and hidden Markov models are suitable for this purpose.

4.2 Definition of Basic Concepts in TCM Systems

4.2.1 Feature

Feature is a quantitative description of a state of interest. It is usually arranged in the form of a feature vector as:

$$y = [y_1, ... y_n]^T$$

where y_1, y_2, . . . , y_n are the features. Depending on the measurements of the class, features can be either discrete numbers or real continuous values. The requirement on features is that the features can reflect the characteristics of desired states and differ from those of other states to largest extent.

4.2.2 State

State (or class) is a set of patterns that share some common properties. The feature vectors of the same type of class will naturally form one set. Due to the diversity of the classes, the features extracted from the same type of classes are seldom identical. This can be interpreted as clusters of points in an *n*-dimensional space, which are called distributions of states. Since the purpose of pattern recognition is to classify these features, the distributions of classes are desired to be separable and not empty. Suppose we have *K* classes, in a mathematical form, the requirement is:

$$\omega_k \neq \phi \quad k = 1, ... K; \tag{4.3}$$

and

$$\omega_k \bigcap \omega_l = \phi \quad k \neq l \in \{1, ... K\} \tag{4.4}$$

4.2.3 Classifier

Classifier is a series of input-output functions $g_i(x, \Lambda_j)$, i = 1, . . . ,*K,* which are *discriminant* between states. In a discriminant function $g(x, \Lambda)$, x is the input vector and Λ is the parameter set of the class. Each discriminant function will output a value. Based on these values, the classifier then assigns x to one of the classes following the decision rule:

$$x \in Class\ i \quad if \quad g_i(x, \Lambda_i) = \max_{for\ all\ j \in K} g_i(x, \Lambda_j) \tag{4.5}$$

There are various classifiers developed in pattern recognition, based on the classification criterion used in the discriminant functions. Classifiers can be grouped into *Bayesian classifier, likelihood classifier* and *distance classifier* [81].

4.3 Data Flow in TCM System

Pattern recognition involves mathematical and technical aspects of classifying different objects through their observable information, such as grey levels of pixels for an image, energy levels in frequency domain for a waveform, and the percentage of certain contents in a product. The objective of a TCM system can be achieved in a three-step procedure, as shown in Figure 4.1.

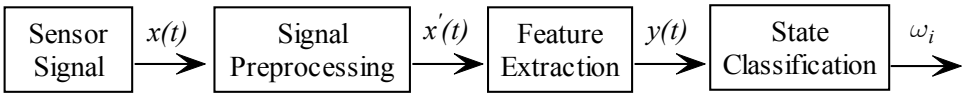


Figure 4.1: Data flow in TCM system

The first step is signal preprocessing. The sensor signal $x(t)$ is firstly pre-processed to normalize the data and remove the noise. The purpose of signal *pre-processing* is to prepare the data for feature extraction.

The next step is feature extraction. The purpose of feature extraction is to make the input data more suitable for pattern classifier and/or reduce the dimensionality of the input data vectors. Information relevant to pattern classification is extracted from $x(t)$ to a feature vector $y(t)$ by a feature extractor. In general, the components of the signal that distinguish the various tool wear classes will be hidden in features that characterize the normal operating condition of the structure, particularly when the tool wear is not severe yet. The task of feature extraction is to enhance the characteristics of the various tool wear classes and suppress or filter off the normal background.

The final stage is state classification. The classifiers decide the tool wear state on the basis of the given feature vector based on a classification measure. Feature vector y is assigned to one of the K classes, $\omega_1,\omega_2, \ldots \omega_K$, by the classifier based on a certain type of classification criteria, such as distance, likelihood and Bayesian, over class models.

4.4 The System architecture of micro-milling TCM

Key components of the proposed and developed architecture and the methodology are shown in figure 4.2. Essentially, ICA is applied to the signal denoising and features are extracted from the most discriminant wavelet features of the denoised force. These features are then used to model different tool condition with the hidden Markov models. The trained HMMs are then used for tool state estimation.

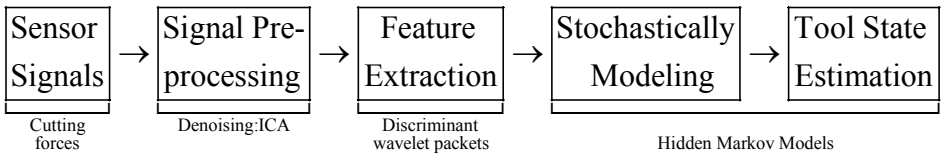


Figure 4.2 System architecture

4.4.1 Signal Pre-Processing

The first stage is signal pre-processing. The purpose of *signal pre-processing* is to prepare the data for feature extraction, which will be discussed in details in later sections. The pre-processing stage basically involves two tasks.

The first is cutting force denoising. In micro-milling, the comparatively small cutting force signal is prone to contamination by relatively large noises, and as a result it is important to denoise the force signal before further processing it. In the ICA approach, noise is taken as a source and sensor signals as an instantaneous mixture of sources (both noise and force). The signal denoising process is treated as blind source separation (BSS), which overcomes the traditional Gaussian noise assumption.

The second task is force normalization. For the real force signal, we are confronted with features vectors that vary within different dynamic ranges because of different

cutting conditions. For example, cutting forces are much larger when the depth of cut is higher. Thus, features with large values may have a larger influence in the discriminant function than features with small values, although this does not necessarily reflect their discriminability of the tool states. This problem is overcome by normalizing the features so that their values lie within similar ranges.

4.4.2 Feature Extraction

The task of feature extraction is to identify and enhance the characteristics of the various tool wear classes from the suitably processed sensor signals. Cutting forces are decomposed by the wavelet packet into five levels with 32 wavelet packet coefficients, and the features are represented by the mean energy of the respective packet node. The features from both wavelet packets and from time domain, i.e. mean, standard deviation, skew, kurtosis and dynamic component, are combined to form a 37-dimension feature vectors.

Furthermore, the dimensionality of parameter vectors is normally very high and needs to be reduced for the sake of less computational cost and system complexity. A feature selection method is used to generate feature vectors by removing one measurement at a time and maintaining the highest value based on some performance indices. There are numerous features that contain redundant information or less sensitive to tool state discrimination. These features need to be eliminated for reduced computation and more robust modeling with HMMs. Fisher's linear discriminant analysis (FDA) is adapted for this purpose. In the discriminant selection, features are chosen to maximize class separation and are ranked by their separation ability between different classes.

4.4.3 Tool Wear State Classification

The final stage is state classification. Continuous Hidden Markov models (HMMs) are adapted for stochastic modeling of the tool wear process in micro-milling, and estimation of the tool wear state based on the output probability of the HMMs given the cutting force features. In the framework of HMMs, the most discriminant features of the cutting forces are selected from both time and wavelet domain first, and then these features are iteratively learned by the HMMs to map the relationships between

force features and tool wear states. Different tool wear states are modeled as separate HMMs. The tool wear state is then classified by the HMM that has the maximum probability to indicate the test features. The framework of HMMs for TCM is illustrated in figure 4.3. The features from both time and wavelet domain are extracted and selected to train the HMMs. A HMM is built for each tool state for classification, and each HMM is trained with its own feature set. Once the models are trained, the $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is obtained that represent different tool wear states. In the tool state recognition state, the extracted features from an unknown state are extracted and input to the HMMs to match the trained states.

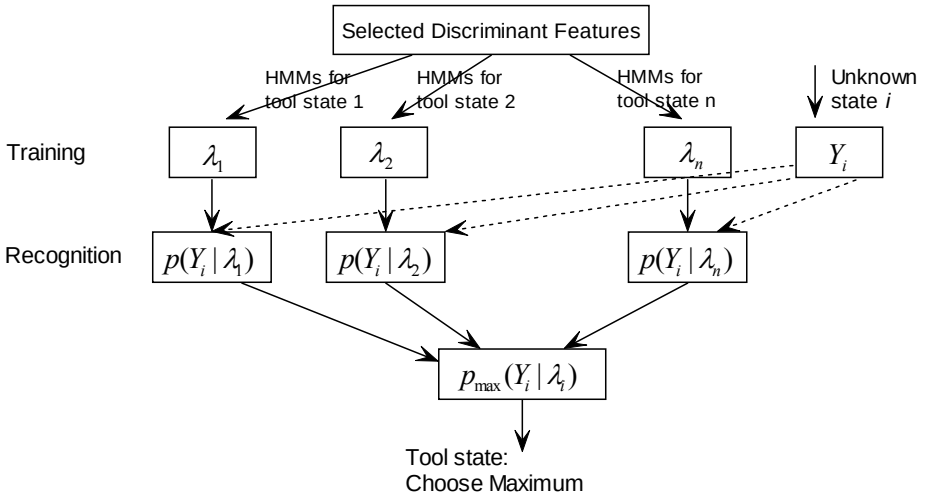


Figure 4.3 Framework of Hidden Markov Models for tool wear classification

Chapter 5
Cutting Force Denoising in Micro-Milling Tool Condition Monitoring

5.1. Introduction

As discussed earlier, cutting forces are comparatively very weak in micromachining and the signal-to-noise ratio (SNR) is very low (about 10 dB). An effective noise removal procedure is an essential first step towards accurate estimation of the tool wear state in micromachining.

Independent component analysis (ICA) is a recently developed method for blind source separation. ICA intends to recover the sources using only the observed signals and on the general assumption that the expected sources are mutually independent. ICA has found wide applications in biomedical systems, speech separation and array processing. Lately it was applied to signals generated by mechanical components and systems. Gelle and Colas [62] introduced the use of blind source separation as a pre-processing approach to improve vibration monitoring, and compared the application in time and frequency domain. Roan et al. [164] applied the Infomax ICA [21] to gear vibration analysis, and showed that the algorithm was able to localize defects in a gear tooth during the learning process. Serviere and Fabry [172] assumed convolute source mixtures and proposed principal component analysis (PCA) based on fourth order cumulants for removal of noisy harmonics. This method is in fact an ICA approach. Li, at al, [110] applied an independent component analysis (ICA) model to identify the engine noise sources and demonstrated that ICA can separate low-level sources effectively. Tian at al. [194] separated the gearbox vibration signals with fixed-point ICA [75] [191] to different frequency bins and used wavelet for the fault diagnosis. Their results showed that the combination of ICA and wavelet was better than using wavelet alone.

However, little has been investigated with the application of the ICA for signal denoising and to the analysis of signals such as those from cutting forces. This chapter introduces a FastICA algorithm as a preprocessor to provide noise free forces for later correlation to tool flank wear. The chapter identified that there exist both Gaussian and non-Gaussian noises. The chapter also assumes that the noises are sources and treat the denoising process as blind source separation. We apply the FastICA for these blind sources separation and then discard the separated noise components. The BSS process is treated as signal denoising in this approach.

The chapter is organized as follows. Section 5.2 introduces the general principles and properties of ICA and the FastICA algorithm. For the purpose of comparing the ICA approach with other technique, the general theory and properties of wavelet thresholding are introduced in Section 5.3. Section 5.4 illustrates the denoising results in the time-frequency domain, by comparing the proposed FastICA and wavelet thresholding to the micro milling force de-noising. Section 5.5 provides analytical discussion on the reason why FastICA is superior to wavelet in the presence of super-Gaussian noise.

5.2 Identification of Noise in Micro-milling

Experiments were conducted to illustrate the application of the FastICA algorithm for separating micro-milling forces. The machine used in the experiments is MAKINO V55 vertical milling machine driven by a 22 kW spindle drive motor. The cutting force was measured with Kistler 9256A dynamometer 3-channel dynamometer mounted under the work piece. Tool wear is measured using the Olympus Toolmakers microscope, 213 times enlargement. The cutting force output was recorded on a Sony digital tape recorder.

Figure 5.1 shows a typical cutting force and noise in the cutting process. From the figure, it can be deduced that the signal to noise ration is very low. In this chapter, the signal to noise ratio (SNR) is roughly estimated as:

$$SNR = 10 \times \log_{10} \frac{y_i^2}{x_i^2} = 10.2 \text{ dB} \qquad (5.1)$$

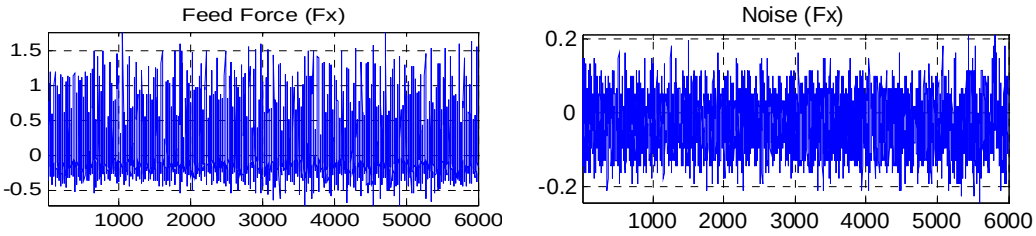


Figure 5.1 Feed force and feed noise SNR

As mentioned in section 5.1, we assume that all the noises are sources. There are three sources: one cutting force source, one environmental noise source and one machining noise source. According to the Center Limit Theory, we can assume that the environmental noise is Gaussian since the samples are very large. As we assume in the ICA, the three sources are linear mixture, that is, sensor forces (observations) are from the weighed linear sum of the three sources, with no time lag and convolution involved. The reference noise is collected during the recess of the machine tool. For the noise in Figure 5.1, we estimated their statistics in table 5.1. The noise is of the normalized kurtosis >0. It is super-Gaussian, with a longer tail than Gaussian distribution (See also Figure 5.15).

Table 5.1 the statistics of reference noise

Statistics	Function	Value
Mean	$\mu = E[x(t)]$	-0.1093
Variance	$\sigma^2 = E[(x(t)-\mu)^2]$	0.2631
Skew	$Skew = E[(x(t)-\mu)^3/\sigma^3$	0.0004
Kurtosis	$Kurtosis = E[(x(t)-\mu)^4/\sigma^4 - 3$	1.2621

5.3 Independent Component Analysis

5.3.1 Motivation: PCA inadequacy

PCA is widely used for condition monitoring for its excellent ability to handle high-dimensional, noisy, and highly correlated data. By applying classifier to the lower-dimensional features produced by PCA, faults can be detected and diagnosed with greater proficiency.

The basic idea in PCA is to find the components $S_1, S_2, \ldots S_n$ that explain the maximum amount of variance of random data X by n linearly orthogonal transformed components. Define the direction of the first principal component w_1, by

$$w_1 = \arg\max E\{w^T x)^2\}, \ \|w\| = 1 \qquad (5.2)$$

Thus the first principal component is the projection along the direction in which the variance of the projection is maximized. Having determined the first $k-1$ principal components, the k-th principal component is determined as the principal component of the residual:

$$w_k = \arg\max E\{[w^T (x - \sum_{i=1}^{k} w_i w_i^T x)^2]\}, \ \|w\| = 1 \qquad (5.3)$$

The i principal component (PC) is then given by $s_i = w_i^T x$. The w_i are the correspond to the n largest eigenvalues of the covariance matric.

PCA is based on second-order statistics. The Second Order Statistics are not sufficient to separate the sources [28] [172]. As shown in figure 5.2, the directions of PC1 and PC2 are not along those of the desired signal independent components (IC1, IC2), but towards the maximum variances (second order). This is because PCA only concerns decorrelation, which is weaker than ICA. ICA and PCA only coincide in the case of Gaussian signals, because the higher-order statistics of Gaussian signals are zero and uncorrelation means independence; thus their second-order statistics define the full characteristic of the signals. However, this is not the case in machining forces, as will be shown in section 5.4, as the forces are not Gaussian. This therefore indicates that independent component analysis will be potentially more appropriate and effective.

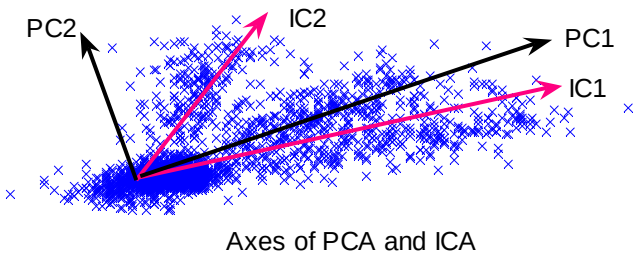


Figure 5.2 PCA vs ICA

5.3.2 Basic Model of Independent Component Analysis

In general independent component analysis is usually referred as a blind source separation method. ICA attempts to recover the hidden source signals from a number of mixed observations from sensors, assuming that the sources are independent with some priori information about their distribution. The problem was addressed by Comon [36], Bell and Sejnowski [21], Cardoso [28], Hyvarinen and Oja [75] [76]. ICA is widely applied to EMG/EEG signal analysis [75], natural image processing [75] [76] and audio/speech signal processing [76].

Random variables $s_i (i = 1, 2, ...m)$ are independent if their joint distribution is a product of their marginal:

$$p(s_1, s_2, ...s_m) = p_1(s_1) p_2(s_2) ... p_m(s_m) = \prod_{i=1}^{m} p_i(s_i) \qquad (5.4)$$

In general, *m* mixed signals $x_i(k)$ $(i = 1, 2, ...m)$ are linear combinations of *n* (typically $m>n$) unknown mutually statistically independent, source signals $s_i(k)$ $(i = 1, 2, ...n)$, and noise $v_i(k)$ $(i = 1, 2, ...m)$ (Fig.5.3). This can be expressed as

$$x_i(k) \ (i = 1, 2, ...) = \sum_{j=1}^{n} w_{ij} s_j(k) + v_i(k) \qquad (5.5)$$

or in the matrix notation

$$X(k) = WS(k) + V(k) \qquad (5.6)$$

where $X(k) = [x_1(k), x_2(k), ...x_m(k)]^T$ is a vector of sensor signals,

$S(k) = [s_1(k), s_2(k), ...s_n(k)]^T$ is a vector of sources,

$V(k) = [v_1(k), v_2(k), ...v_m(k)]^T$ is a vector of additive noise, and

W is an unknown full rank $m \times n$ mixing matrix.

This instantaneous mixing process is illustrated in figure 5.3.

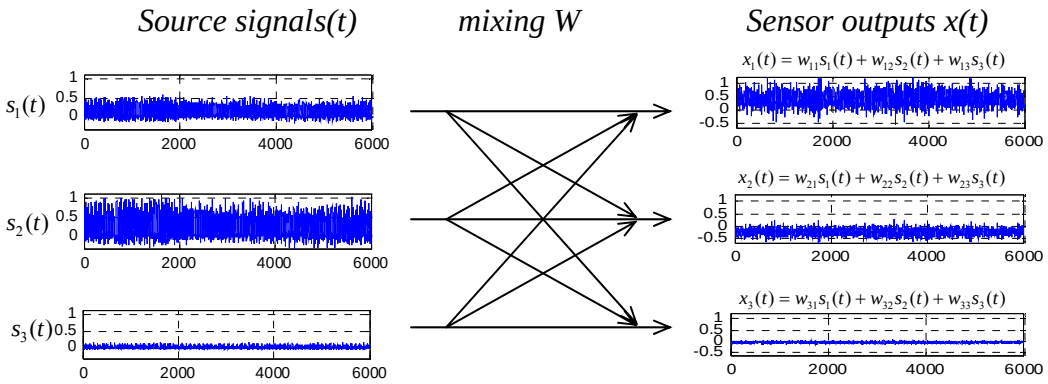


Figure 5.3 Linear mixing of sources

Our aim is to find some methods to reverse this mixing process (un-mixing) and find the original sources. Assume that there is a matrix H and,

$$H\mathbf{x}(t) = H[W\mathbf{s}(t)] = (HW)\mathbf{s}(t) \tag{5.7}$$

$$If \quad HW = I, \ i.e.\ H = W^{-1}, \qquad then\ H\mathbf{x}(t) = I\mathbf{s}(t) = \mathbf{s}(t) \tag{5.8}$$

In this way, if we can find some un-mixing matrix $\boldsymbol{H}$, we will recover the sources $\mathbf{s}(t)$. This whole separation process is illustrated as figure 5.4.

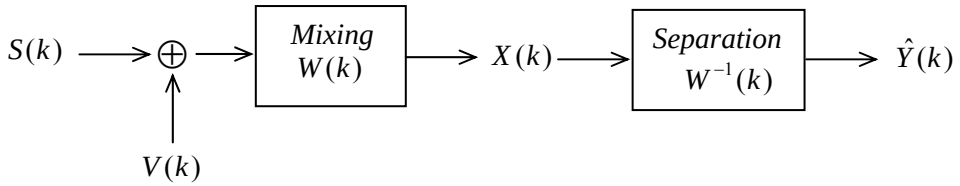


Figure 5.4 General model of ICA: $X(k)$ observations, $Y(k)$ estimated sources, $S(k)$ sources

Generally the unmixing matrix can be found by virtue of the non-Gaussian nature of the underlying sources. The classical measure of nonGaussianity is kurtosis, or the fourth-order cumulant. The kurtosis of $\boldsymbol{y}$ is classically defined by

$$Kurt(X) = E[x(t)]^4 / \sigma^4 - 3 \tag{5.9}$$

Actually, the kurtosis is zero for a Gaussian random variable, and nonzero for nonGaussian random variables.

The majority of past research efforts have been devoted to only the noiseless case, where $V(k) = 0$, as the estimation of a separating (un-mixing) matrix $\boldsymbol{H}$ and a mixing

matrix ***W*** in the presence of noise is rather difficult because it is an ill-conditioned problem, and only solvable in certain "sparse" cases [220]. We simplify this problem by treating the noise as source signals, and as a result we deal with the noiseless case but more sources. This assumption is probable and solvable in our case because we have three force components that are actually from only one force source, or say we have more sensors than sources, which will be discussed in section 5.5.

5.3.3 FastICA: Negentropy as a Measure of NonGaussianity

However, the kurtosis is sensitive to outliers [28][75] and is not a robust measure of nonGaussianity. In this approach, we apply the FastICA and measure the nonGaussianity with negentropy. Take $f(y)$ as the probability density function (pdf) of random variable y, the entropy of y is then defined as:

$$H(Y) = -\int f(y)\log(f(y))dy \tag{5.10}$$

A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance. This means that negentropy could be used as a measure of nonGaussianity. Negentropy is defined by:

$$J(Y) = H(Y_{gauss}) - H(Y) \tag{5.11}$$

where $\boldsymbol{Y_{Gaussian}}$ is a multivariate normal with the same covariance matrix as ***Y***. Negentropy measures how far a signal from its comparative Gaussian signal. However, it is hard to compute ***J(Y)*** directly. ***J(Y)*** may be approximated by:

$$J(Y) \approx k[E(G(Y)) - E(G(\nu))]^2, \tag{5.12}$$

For some $\boldsymbol{k \geq 0}$, $\boldsymbol{\nu \sim N(0,1)}$ and ***Y*** is standardized.

Indeed, taking $\boldsymbol{G(y) = y^4}$, one then obtains a kurtosis-based approximation. ***G*** here is a non-quadratic function and is picked to avoid the problems of kurtosis with outliers. Here we choose ***G(u)=tanh(u)***.We obtain approximations of negentropy that are much better than kurtosis.

To estimate several independent components, we need to run the one-unit FastICA algorithm using several units (e.g. neurons) with weight vectors $\boldsymbol{w_1, \ldots w_n}$. To prevent different vectors from converging to the same maxima we decorrelate the outputs

$w_1^T x,...,w_n^T x$ by deflation after each iteration. The ***p-th*** extraction processing unit extracts a source signal from inputs that are linear mixtures of the remaining source signals. The ***p-th*** deflation processing unit then removes the newly extracted signal from the mixtures and feeds the resulting outputs to the next ***(p+1)-th*** extraction processing unit. The procedure can be continued until all the estimated source signals are recovered. This process is implemented as:

$$\text{1. Let } w_{p+1} = w_{p+1} - \sum_{j=1}^{p} w_{p+1}^T w_j w_j$$

$$\text{2. Let } w_{p+1} = w_{p+1} / \sqrt{w_{p+1}^T w_{p+1}} \qquad (5.13)$$

Table 5.2 illustrates the implementation of the proposed algorithm.

Table 5.2 FastICA algorithm

1. Center data to zero mean and whiten the data to give z .
2. Choose *m*, the number of ICs to estimate. Set counter $p \leftarrow 1$.
3. Choose an initial value of unit norm for w_p randomly
4. Let $w_p \leftarrow E\{zg(w_p^T z)\} - E\{g'(w_p^T z)\}w$, where g is defined as,

$G(u) = \tanh(u)$

5. Do the following orthogonalization:

$$w_p \leftarrow w_p - \sum_{j=1}^{p-1} (w_p^T w_j) w_j$$

6. Let $w_p \leftarrow w_p / \|w_p\|$
7. If w_p has not converged, go back to 4
8. Set $p \leftarrow p+1$. If $p \le m$, go back to 3

Figure 5.5 illustrates the implementation of the proposed algorithm. The $j\text{-}th$ extraction processing unit extracts a source signal from inputs that are linear mixture of the remaining source signals yet to be extracted. The $j\text{-}th$ deflation processing unit then removes the newly extracted signal from the mixtures and feeds the resulting outputs to the next $(j+1)\text{-}th$ extraction processing unit. The procedure can be continued until all the estimated source signals are recovered. We can achieve this by

setting a threshold for the amplitude of each signal x_{j+1}. Fig. 5.5 Block diagrams illustrating implementation of sequential extraction and deflation principles.

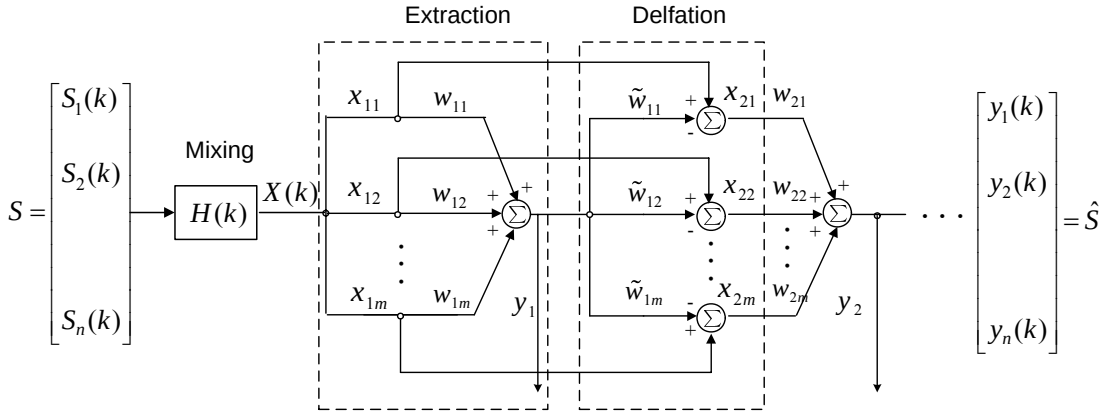


Fig. 5.5 Block diagrams of implementation of sequential extraction and deflation

5.4 Source Separation in Micro-milling

Now apply the FastICA to the cutting forces separation. Three sources are expected: one force source, one non-Gaussian source for machining noise and one Gaussian source of environmental noise. Noises are treated as sources here. The sensor outputs and their corresponding power spectrum are illustrated in Figure 5.6. The forces are sampled with frequency 6,000 Hz. As can be seen from the figure, the low frequency components have the highest peaks all along the time axis and as a result the desired harmonics ***nFr*** are immersed.

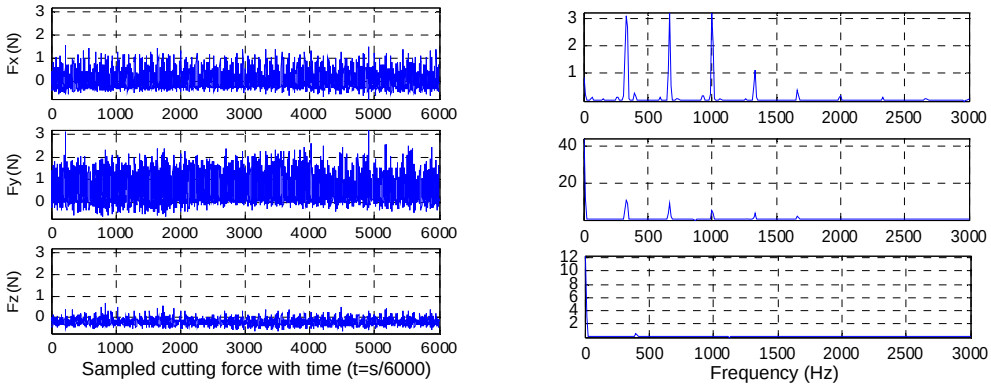


Figure 5.6 the sensor output and their corresponding power spectrum

And now we apply the FastICA. The whitened forces are illustrated in Figure 5.7, and the reconstructed sources and their corresponding power spectrum are illustrated in

Figure 5.8. It is clear that reconstructed force peaks at the first five harmonics, while the machining noise peaks at low frequency and a wide range of harmonics, and the environmental noise peaks at very low frequency.

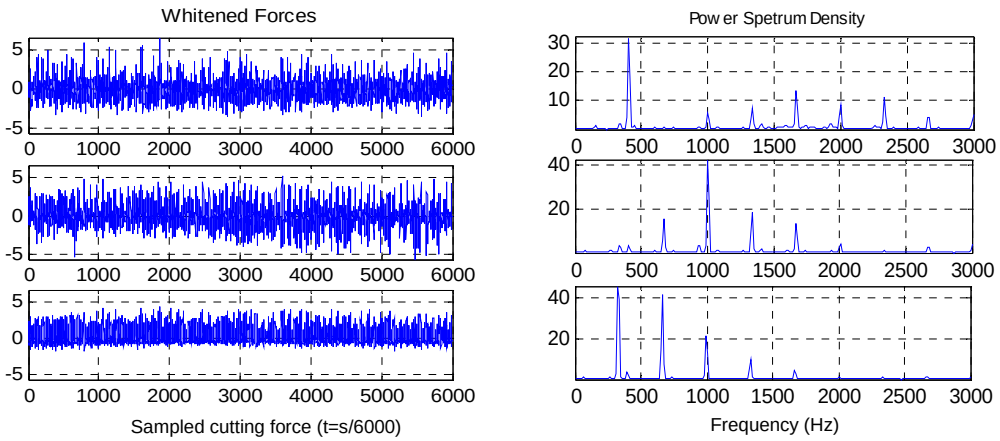


Figure 5.7 the whitened forces and their corresponding power spectrum

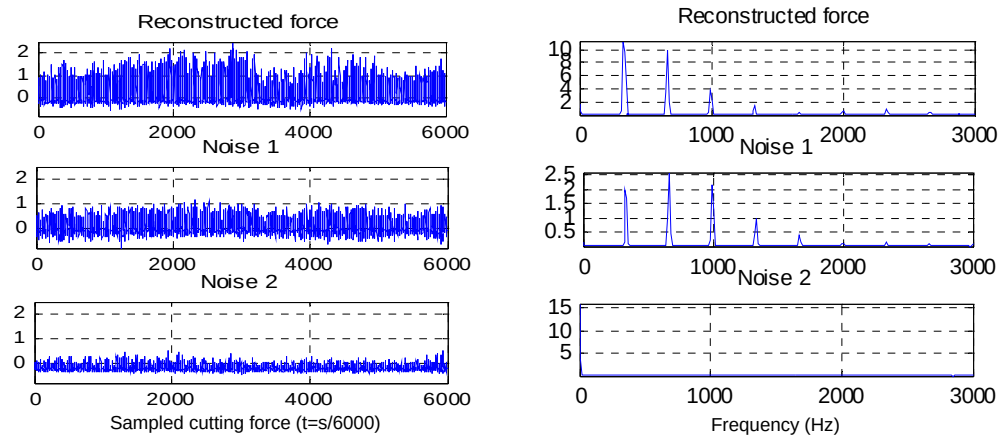


Figure 5.8 Cutting forces reconstructed

We now apply the Short Time Fourier Transform (STFT) to demonstrate the full pass results. The interest is primarily in the few frequency bins of harmonics nF_r. With the Short Time Fourier Transform,

$$STFT_x(t,\Omega) =< x(\tau), g_{t,\Omega}(\tau) >= \int x(\tau) g^*(t-\tau) e^{-j\Omega\tau} d\tau \qquad (5.24)$$

where $g(\tau)$ is window function and sliding along the time axis.

All the reconstructed force are added together and illustrated by spectrogram in Figure 5.9. It is observed that after 120 second the *1st* , *2nd*, and *3rd* harmonics of the feed force increased dramatically, corresponding to the sharp increase of tool flank wear in figure 5.10. It is observed that after 22.00 minutes the power spectrum of feed force

increased largely, which corresponds to the dramatically increasing of measured tool flank wear.

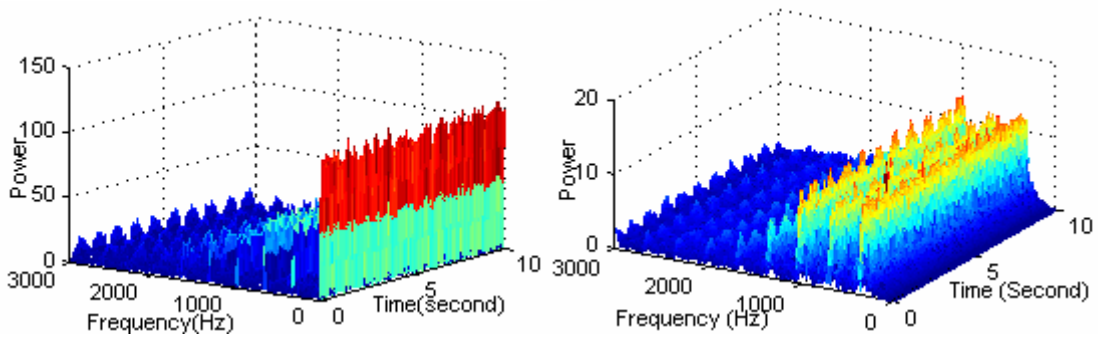


Figure 5.9 a) Force spectrogram before ICA Figure 5.9 b) Force spectrogram after ICA

Figure 5.9 Force spectrogram before and after ICA

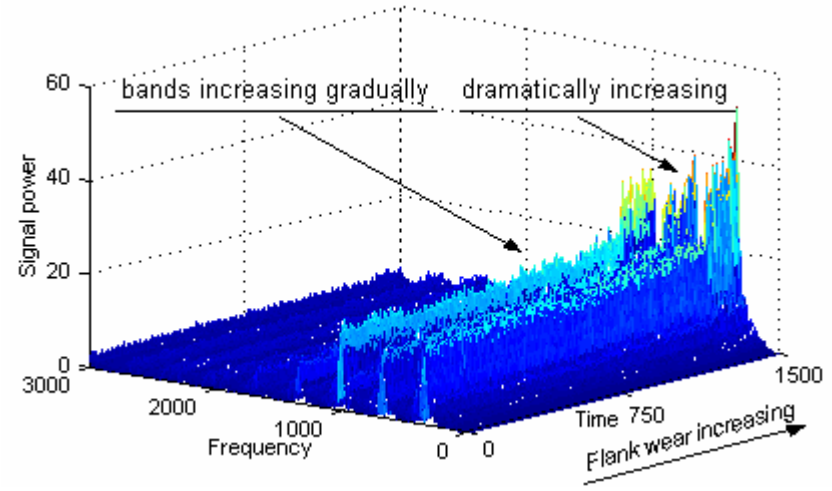


Figure 5.10 Tool life with reconstructed feed force

To show the effectiveness of FastICA in the removal the noise, we compare it with wavelet thresholding for denoising the same forces. Here we choose soft thresholding with feed force ***Fx*** (figure 5.11), as feed force is more correlated to the flank wear in these experiments (Appendix A). The wavelet thresholding result is illustrated in figure 5.12. There is little difference between the two figures. This phenomenon due to the poor performance of wavelet thresholding on non-Gaussian noise, it will be discussed in Section 5.5. The same passes were added together to show their overall denoising performance in figure 5.13. Comparing the figure 5.10 and figure 5.13, we can see figure 5.10 shows better diagnostic information since harmonics stand out and the first three harmonics increasing dramatically according to the fast increase of flank wear.

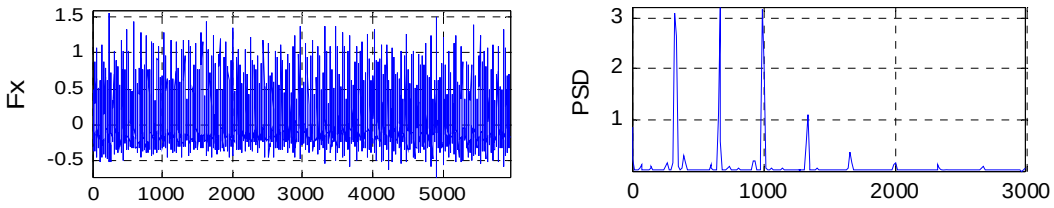


Figure 5.11 Feed force and their corresponding power spectrum density

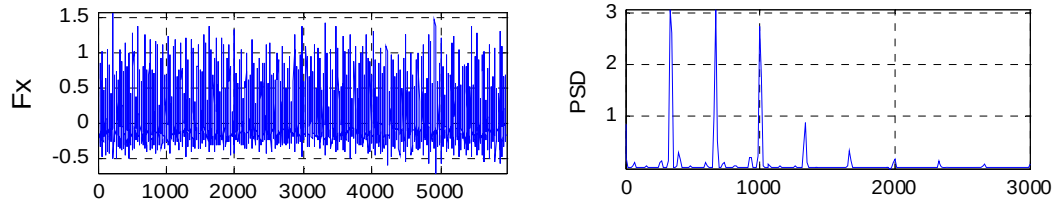


Figure 5.12 Wavelet denoised force and their corresponding power spectrum density

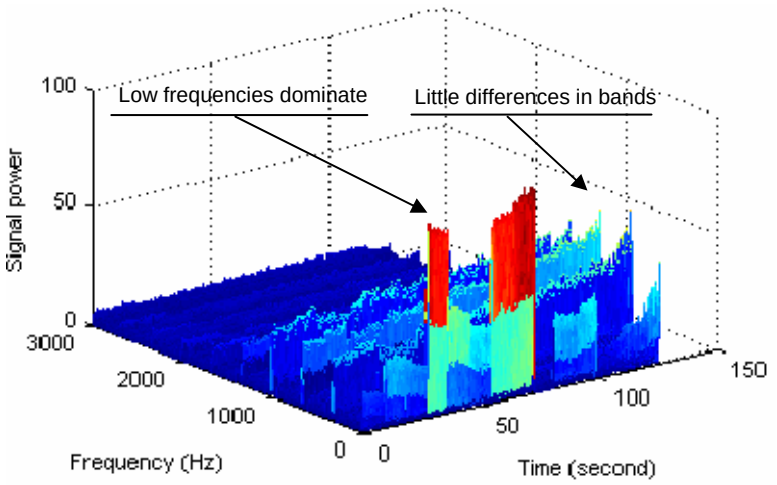


Figure 5.13 Tool wear with wavelet denoised force

5.5 Discussion

5.5.1 ICA Solvability Analysis

According to the Center Limit Theory, any random process obeys Gaussian distribution when it approaches infinity. This means that a sum of several variables typically has a distribution that is closer to Gaussian than any of the original random variables. Since the samples taken are very large, we assume that the random noise is

Gaussian. To estimate one of the independent components, we consider a linear combination of the x_i. Let us denote this by, $y = w^T x = \sum_i w_i x_i$ where $\boldsymbol{w}$ is a vector to be determined. If $\boldsymbol{w}$ were one of the rows of the inverse of **A**, this linear combination would actually be equal to one of the independent components. In practice, we cannot determine such a $\boldsymbol{w}$ exactly, because we have no knowledge of matrix **A**, but we can find an estimator that gives a good approximation.

To see how this leads to the basic principle of ICA estimation, let us make a change of variables, defining $z = H^T w$. Then we have $y = w^T x = w^T Hs = G^T s$. y is thus a linear combination of $\boldsymbol{s_i}$, with weights given by $\boldsymbol{G_i}$. Since a sum of even two independent random variables is more Gaussian than the original variables, $G^T s$ is more Gaussian than any of the $\boldsymbol{s_i}$ and becomes least Gaussian when it in fact equals to one of the $\boldsymbol{s_i}$. In this case, obviously only one of the elements $\boldsymbol{G_i}$ of $\boldsymbol{G}$ is nonzero.

Therefore, we could take $\boldsymbol{w}$ as a vector that ***maximizes the nonGaussianity*** of $w^T x$. Such a vector would necessarily correspond (in the transformed coordinate system) to a vector which has only one nonzero component. This means that $w^T x = G^T s$ equals one of the independent components. Maximizing the nonGaussianity of $w^T x$ thus gives us one of the independent components.

5.5.2 Wavelet Thresholding Assessment

The basic property of wavelet decomposition that leads to noise reduction is sparsity. For noise reduction purposes, sparsity of the coefficients is utilized where sparsely distributed coefficients having large amplitudes, carry much of the signal information while in contrast, small coefficients may be considered insignificant and belong to the noise content of the signal and thus may be removed by thresholding. However, there be difficulty in the thresholding when dealing with the super-Gaussian noise (i.e. noise having heavy tails and outliers), and this is just the case in micro-milling.

A central question in many thresholding procedures is the selection of the threshold. Donoho and Johnston [43], [44] proposed the threshold that comes close to the minimax threshold: the threshold that minimizes the worst case mean square error in a typical function space. But they pay more attention to smoothness than to minimizing

the mean square error. The threshold is based on the white Gaussian noise, which is too strong an assumption in practical situation (eg. Micro-milling).

5.5.2.1 Gaussian Noise

Theorem: An orthornormal wavelet transform of stationary and white Gaussian signal is still stationary and white Gaussian signal with the same statistical properties.

Proof:

For any signal contaminated with Gaussian noise,

$$y_i = f(x_i) + e_i, \text{ where } e_i \sim N(\mu_i, \sigma^2) \text{ or } \quad Y = f + \varepsilon \tag{5.15}$$

Wavelet transform of Y:

$$W(Y) = W(f + \varepsilon) = Wf + W\varepsilon \tag{5.16}$$

is a linear transformation, we consider the noise term $W\varepsilon$ only.

$\mu_i = 0,$ then it follows that $\mu_\varepsilon = E(W\varepsilon) = WE(\mu_i) = 0;$
$\Sigma_i = \sigma^2 I_N,$ *and* $\sin ce\ WW^T = I_N$ due to orthonomality,
$\Sigma_\varepsilon = W\Sigma_i W^T = \sigma^2 W I_N W^T = \sigma^2 WW^T = \sigma^2 I_N$

Since the Gaussianity of e implies the Gaussianity of ε, we can conclude that the statistical properties of e and ε are identical. Therefore the noise has the same characteristics both in the original domain and the wavelet domain.

If the noise is white Gaussian, it affects all the coefficients equally (Figure 5.14). The noise is spread out evenly over all coefficients, and since the power is limited, the noisy coefficients will be very small in all bands. According to the ***sparse*** representation by wavelet, signal ***f(x$_i$)*** is sparse and has few large magnitude coefficients and all the others are very small (close to zero) in wavelet domain**.** These few large coefficients describe fully of the signal, and the small coefficients are noise. Thus we keep the bulk of important coefficients, and remove the small ones that by thresholding, and since only a few remain, most noise is removed.

The residue of the denoised force is shown in figure 5.14. From figure 5.14(c) the PSD harmonics distribute nearly evenly in all frequency bands; this explains why it can not be separated even in the low-frequency noise environmental. In figure 5.14(d), the

autocorrelation is close to zero; this means that the estimated noise is purely random, not as shown in figure 5.14 d). Figure 5.14(b) shows that it is typically a Gaussian noise.

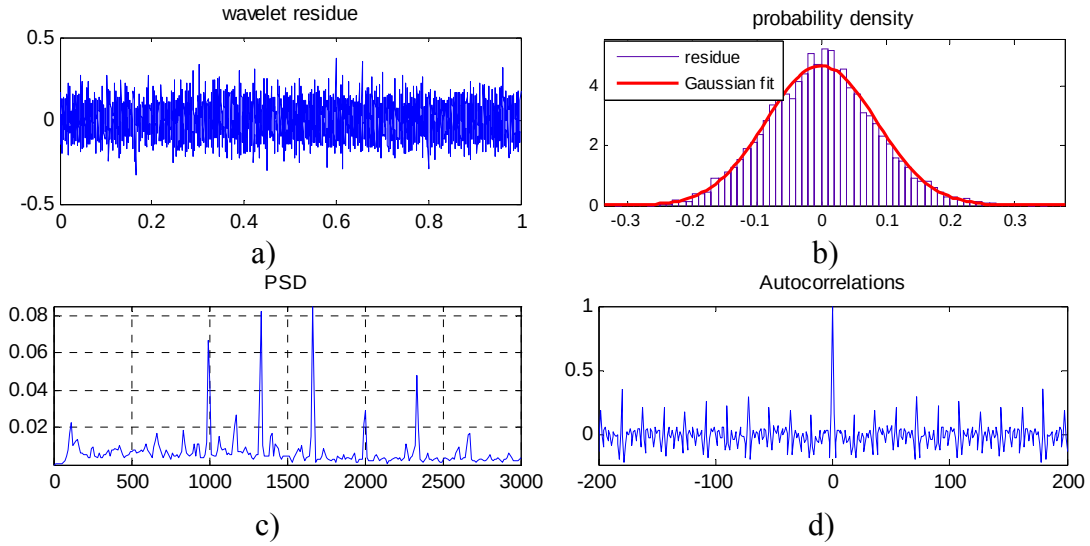


Figure 5.14 Illustrative of the of wavelet residue: a) residue b) the corresponding power spectrum density, c) noise distribution compared with Gaussian distribution, d) the autocorrelation coefficients

5.5.2.2 Non-Gaussian Noise

Approaches of wavelet-based denoising have generally relied on the assumption on Gaussian noise and are therefore sensitive to outliers, i.e., to noise distributions whose tails are heavier than the Gaussian distribution, such as Laplacian and α-stable distributions [12] [16].

It has been shown by Ruggeri and Vidakovic [166] that the threshold is sensitive to the prior distribution and it should be careful in choosing the parameter values. A variety of models of interest to engineering includes Laplacian distribution and Cauchy distribution whose distributions have heavier tails than the Gaussian. Examples of these noises include atmospheric noise, cellular communication, underwater acoustics, and moving traffic [12] [16].

For non-Gaussian noise it rather complicated to compute the distribution of the noise in the upper levels of the transformed signal. If this noise is heavy tailed than the thresholds, the upper levels are much higher than necessary. For example, [82] [166]

dealt with a random variable with a Laplacian distribution. It is expect that the optimal thresholds for higher level wavelet coefficients are much smaller than the optimal thresholds for the Laplacian distribution, as in our case (Figure 5.15). In the figure 5.15, the power spectrum density (PSD) and autocorrelation show that the noise is not white. The PSD harmonics peak at the characteristic frequencies (*nFr*) and the autocorrelation is very high. Typically it resembles a component of cutting force.

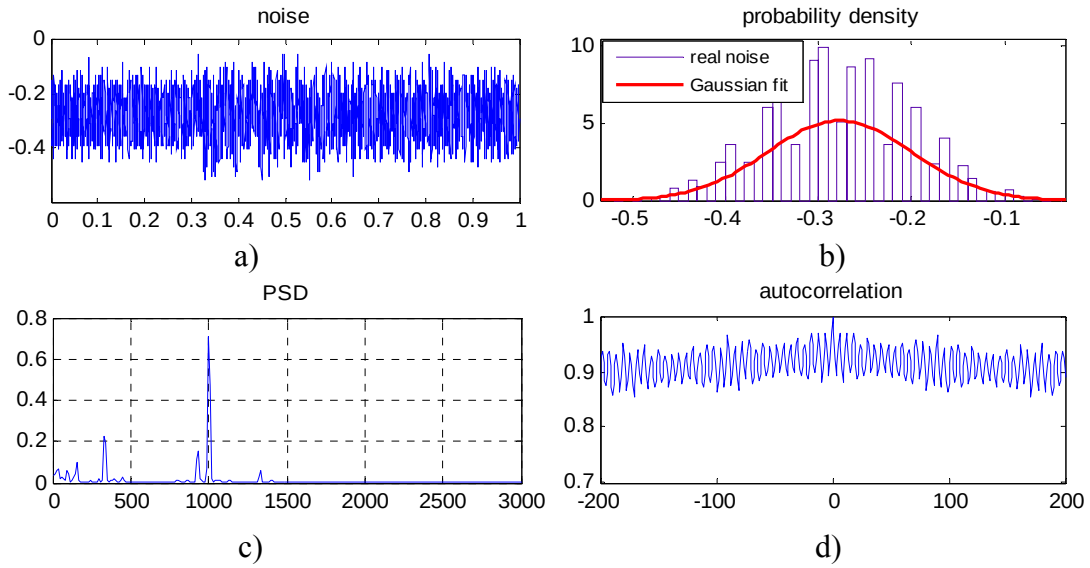


Figure 5.15 Illustrative of the statistics of noise: a) noise signal, b) noise distribution compared to Gaussian distribution , c) the corresponding power spectrum density, d) the autocorrelation coefficients

5.6 Conclusion

FastICA shows better force denoising performance than wavelet for use in micro milling tool wear monitoring. In micromachining there are machining noises that are correlated to the machining conditions and have wide harmonics and non-Gaussian distribution. FastICA denoises force efficiently by directly separating Gaussian and non-Gaussian noises as signal sources, under the measure of negnentropy. In this process, the non-Gaussian noise or noise harmonics are removed as source separation. Wavelet thresholding estimates too small threshold in this case, and does not work well by discarding smaller coefficients since the non-Gaussian noise coefficients are not small and evenly distributed in certain bands.

Chapter 6
Discriminant Feature Selection for Hidden Markov Models in Micro-milling Tool Wear Classification

6.1 Introduction

For an effective modeling of tool wear state with HMMs, however, good feature representation of different tool state is prerequisite. In TCM some candidate features may be irrelevant and resulted in adding noise to useful information, or redundant with the presence of other similar features. These features do not contribute to the efficient classification of tool conditions. Instead, they contribute to unnecessary computation. Tool wear classification algorithms should consider whether the features are independent and discriminating, and avoid unnecessary computation or hinder the subsequent classification process. A feature selection algorithm to deal with this problem is necessary.

Goebel and Yan [65] carried out a thorough comparison with soft computing methods for feature selection in milling. However, there was no good generalization even 25 features from 5 different schemes were compared, both from time frequency domain. Among the feature dimension reduction approaches, principal component analysis (PCA) is one of the classic methods and widely studied. Ali and Elijah [5] presented a large improvement of classification rate after transforming the entire feature sets with PCA and then choose only the top 6 principal components of the 46-dimentional features for discrimination in TCM. Li and Elbestawi [109] also applied the PCA for the feature dimension reduction for clustering of different tool wear in turning. But as unsupervised learning approaches, PCA may throw away the information required to perform state classification that lies in the small eigenvalue's direction. These directions are treated as noise with PCA procedure. Another popular feature selection scheme is by bayesian approaches. One of the most successful methods is automatic relevance determination (ARD) [129]. It is widely studied in machine learning literature [156] [195] and applied to feature selection [35]. Sun et al [180] investigated

ARD for acoustic emission feature selection under a support vector machine (SVM) for classification. It was able to identify the worst features according to their corresponding ARD parameters and then deleted them. However, ARD finds the relevance of features by optimizing the model marginal likelihood, no within-feature correlation considered. This may also bring the problems of over-fitting.

In this chapter, we focus on the selection of features for HMMs in this chapter, but not on the design of the structure of HMMs. We adapt the Fisher's linear discriminant analysis (LDA) [81] for feature selection that ranks features by their class-discriminant ability. Unlike PCA and ARD for feature selection, the problem of selecting a subset of features in LDA reduces to picking only the most discriminant wavelet packets for tool state representation.

6.2. Wavelet Packet Decomposition of Cutting Forces

Due to the nature of machining processes, the signals are usually time varying and nonstationary. Wavelet analysis overcomes the pitfalls of short time Fourier methods and is adapted for processing nonstationary signals. It has been widely studied to analyze machining signals in TCM such as Wu and Du [210], Pittner and Kamarthi [158], and Tansel et al [186]. It was found that wavelet analysis can capture important features of signal that are sensitive to different tool wear states, but is insensitive to the variation of working conditions and noises. This provides wavelet a powerful and robust tool for TCM.

Wavelet packet decomposition approximates a signal with scaled and translated wavelet packet functions,

$$W_{j,k}^{n}(t) = 2^{-j/2} W^{n}(2^{-j/2} t - k) \tag{6.1}$$

which characterizes the local variation $2^{j}k$ in scale 2^{j}, and n is the oscillation parameter $n = 1, 2, \ldots$. The wavelet packet $W_{j,k}^{n}(t)$ forms orthogonal wavelet packet basis. The wavelet packet coefficients are then produced from this integral:

$$d_{j,k}^{n}(t) = \int W_{j,k}^{n}(t) f(t)\, dt \tag{6.2}$$

Figure 6.1 demonstrates an example of wavelet packet decomposition of cutting force into 5 levels with Daubeches 8 wavelet. There are $N = 2^1 + 2^2 + 2^3 + 2^4 + 2^5 = 62$ (wavelet packets) in this decomposition. There are several algorithms on the selection of the best packets for representation and reconstruction of the signal. What we consider here is not the best signal representation, but best features that characterize the signal from different tool wear level. So we use all the level-5 wavelet packets, but none from the other levels, since this level segments the force into the largest parts (32 wavelet packets) and separates the signal's low frequency and high frequency furthest apart. The frequency bands are mutually disconnected and represent different information of the cutting force.

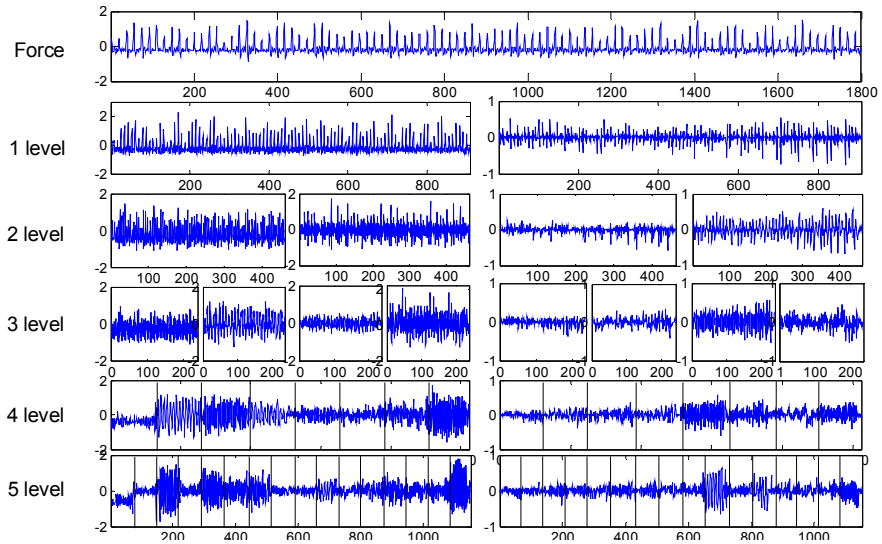


Figure 6.1 a) 5-level wavelet packet decomposition of cutting force. The forces are sampling points with sampling frequency 6,000 Hz.

0-3000 (Hz)																
0-1500								1500-3000								
0-750				750-1500				1500-2250				2250-3000				
0-375		375-750		750-1125		1125-1500		1500-1875		1875-2250		2250-2625		2625-3000		
0-188	188-375	375-563	563-750	750-938	938-1125	1125-1313	1313-1500	1500-1688	1688-1875	1875-2063	2063-2250	2250-2438	2438-2625	2625-2813	2813-3000	
0-94	94-188	188-282	282-375	375-469	469-563	563-657	657-750	750-844	844-938	938-1032	1032-1125	1125-1219	1219-1313	1313-1407	1407-1500	
1500-1594	1594-1688	1688-1782	1782-1875	1875-1969	1969-2063	2063-2157	2157-2250	2250-2344	2344-2438	2438-2532	2532-2625	2625-2719	2719-2813	2813-2907	2907-3000	

Figure 6.1 b) frequency bands of 5-level WPD

These wavelet packet coefficients are then chosen to form feature vectors for HMM modeling after the following feature selection.

6.3. Selection of Discriminant Features

Apart from the original intention of the wavelet transform for time-frequency analysis of non-stationary signals, the important and successful application of the wavelet in TCM is fault feature extraction and dimension reduction. These wavelet coefficients usually need to select for robust and effective representation of different tool states in TCM. The extraction approaches are generally based on some measurements to maximize the discriminability or classification among different tool states. Studies were reported by Yen and Lin [212] with FDA, Malhi and Gao [130] with PCA, and Liu and Ling [122] with mutual information for selecting wavelet packets.

6.3.1 Principal Components Analysis for Feature Selection

Principal Components Analysis (PCA) is widely used for machining monitoring for its excellent ability to handle high-dimension, noisy, and highly correlated data [84]. By applying classifier to the lower-dimensional features produced by PCA, faults can be detected and diagnosed with greater proficiency.

The basic idea in PCA is to find the components $s_1, s_2, \ldots s_n$ so that they explain the maximum amount of variance possible by n linearly transformed components. The principal components are then given by $s_i = w_i^T x$. In practice, the computation of the w_i can be simply accomplished using the sample covariance matrix $E[xx^T] = C$. The w_i are the eigenvectors of C that correspond to the n largest eigenvalues of C. This is achieved by eigen-decomposition of the covariance matrix.

Define the direction of the first principal component, say w_1, by

$$w_1 = \arg\max E[w^T x)^2], \ \|w\| = 1 \tag{6.3}$$

where w_1 is of the same dimension m as the random data vector x.

The $k\text{-}th$ principal component is determined as the principal component of the residual:

$$w_k = \arg\max E[w^T (x - \sum_{i=1}^{k} w_i w_i^T x)^2], \ \|w\| = 1 \tag{6.4}$$

We can view the PCA as orthogonal projections of the data onto the subspace spanned by the first *n* largest eigenvectors of the covariance matrix. The directions corresponding to the small eigenvalues are interpreted as imply less information.

It should be noted that the PCA approach for feature dimension reduction discussed above actually transforms the features to another space. It is feature extraction method, not for feature selection, and the transformed features lose the physical nature of the signals as a result, which is not favored in TCM. In this study the feature selection of PCA is based on the original features, as what is reported in [130] [137]. Our approach is based on Jolliffe [84] for feature ranking. It starts with the eigenvector of the smallest eigenvalue of the covariance matrix, and discards the feature with the largest coefficient in that vector, and then proceeds to the next eigenvector and discards the remaining features with the largest coefficient. This process is iteratively continued until all features are ranked.

6.3.2 Automatic Relevance Determination for Feature Selection

Automatic Relevance Determination (ARD) is a single hidden layer neural network (NN) with varying numbers of hidden nodes was used to model the relationship between the inputs and the outputs. The network is trained by Bayesian learning. Consider a data set $D = \{x_i, y_i\}_{i=1}^{N}$ drawn from an unknown probability distribution, where x_i represents a set of inputs and y_i represents a set of outputs. The underlying mapping of $x \rightarrow y$ is to be modeled with the *NN*.

The distribution of output given input vector x can be written as,

$$P(y \mid x, D) = \int P(y \mid x, w) P(w \mid D) dw \tag{6.5}$$

where $P(w \mid D)$ is the posterior distribution of the model parameters.

Following the work of MacKay [129] a Gaussian prior was chosen for the initial values of the weights. Bayesian learning in the NN simplifies to finding the weight vector, *w*, which minimizes the cost function,

$$E(w) = \frac{\beta}{2} \sum_{n=1}^{N} (y_n - f(x_n; w))^2 + \frac{\alpha}{2} \sum_{i=1}^{W} w_i^2 \tag{6.6}$$

where N represents the number of data points in dataset D.

This cost function is minimized by using the scaled conjugate gradient algorithm while α and β are hyperparameters that control the cost function and are re-estimated using the evidence framework.

The hyperparameters are assigned one-to-one with the set of input features to the MLP, where it represents the inverse variance of the prior distribution of the weights from that input. The optimal hyperparameter values found then reflect the relevance of their associated features to the output with Bayesian learning. As the hyperparameter represents the inverse variance of the weights, a small hyperparameter value means large weights are allowed, and we can conclude that the corresponding feature is important. A large hyperparameter value means that the weights are constrained to near zero, and hence the corresponding feature is less important. This adapts the ARD for feature ranking and selection.

6.3.3 Discriminant Analysis for Feature Selection

The Fisher's linear discriminant analysis (FDA) [49] is generally a classification method to find the projection that maximizes the distance between the means of the two classes while minimizing the variance within each class. We adapt this method for the feature selection in this chapter. Instead of finding the projections, we calculate the score of discriminating ability of features for different classes.

Suppose we have K classes, and we have a set of n d-dimensional feature samples $\{x_1,...x_n\}$, n_1 in the subset D_1 labeled C_1 and n_2 in the subset D_2 labeled C_2.

The d-dimensional sample mean vector μ_i and the covariance matrix S_i of class i are given by:

$$\mu_i = \frac{1}{n_i} \sum_{x \in D_i} x \tag{6.7}$$

$$S_i = \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T , \qquad i = 1,...K \tag{6.8}$$

The within-class covariance matrix S_w is given by:

$$S_w = \sum_{i=1}^{K} S_i = \sum_{i=1}^{K} \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T \tag{6.9}$$

Define the overall mean μ as:

$$\mu = \frac{1}{n} \sum_{x} x = \frac{1}{n} \sum_{j=1}^{K} n_i \mu_i \tag{6.10}$$

Define the between-class covariance matrix S_B:

$$S_B = \sum_{i=1}^{K} n_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{6.11}$$

Let $x = (x_1, x_2, ..., x_d)$ be s set of training vectors from the *n*-dimensional input space R^n. The set of vectors $y = (y_1, y_2, ..., y_d)$ is a lower dimensional representation of the input training vector x in the m-dimensional space R^m. The vector x are obtained by the linear orthonormal projection

$$y_i = w_i^T x_i \quad i = 1, 2, ...m \tag{6.12}$$

$$y = W^T x$$

The corresponding mean and covariance matrix of y are defined by:

$$\mu_i' = \frac{1}{n_i} \sum_{y \in D_i} y \tag{6.13}$$

$$S_i' = \sum_{y \in D_i} (y - \mu_i')(x - \mu_i')^T , \qquad i = 1, ...K \tag{6.14}$$

The within-class covariance matrix S_w' is given by:

$$S_w' = \sum_{i=1}^{K} S_i' = \sum_{i=1}^{K} \sum_{y \in D_i} (y - \mu_i')(y - \mu_i')^T \tag{6.15}$$

Define the overall mean μ' as:

$$\mu' = \frac{1}{n} \sum_{y} y = \frac{1}{n} \sum_{j=1}^{K} n_i \mu_i' \tag{6.16}$$

Define the between-class covariance matrix S_B':

$$S'_B = \sum_{i=1}^{K} n_i(\mu'_i - \mu')(\mu'_i - \mu')^T \tag{6.17}$$

It is straightforward to show that:

$$S'_w = W^T S_w W \tag{6.18}$$

and

$$S'_B = W^T S_B W \tag{6.19}$$

The Fisher's Linear Discriminant is then defined as the linear functions $W^T x$ for which the criterion function

$$J(W) = \frac{|S'_B|}{|S'_w|} = \frac{W^T S_B W}{W^T S_w W} \tag{6.20}$$

is maximum.

It can be shown that the solution of (3.18) is that the *i*-th column of an optimal W is generalized eigenvector corresponding to the *i*-th largest eigen value of matrix $S_W^{-1} S_B$. Noting that the inter-class distance and intra-class distance ratio

$$R = \frac{\text{Inter-class distance}}{\text{Intra-class distance}} = \frac{|S_b|}{|S_w|} \tag{6.21}$$

has large values when samples are well centered around their mean within each class, while different classes are well separated.

In this case, it is easy to see that for equiprobable classes $|S_w|$ is proportional to $s_1 + s_2$ and $|S_b|$ is proportional to $(\mu_1 - \mu_2)^2$. Combing S_b and S_w**,** the so-called *Fisher's discriminant ratio (FDR)* [192] results in

$$FDR = \frac{|\mu_1 - \mu_2|^2}{s_1 + s_2} \tag{6.22}$$

For the multiclass case, averaging forms of *FDR* is used:

$$FDR(m) = \sum_{i=1}^{K} \sum_{j \neq i}^{K} \frac{|\mu_i^m - \mu_j^m|^2}{s_i^m + s_j^m} \tag{6.23}$$

where the subscripts ***i, j*** refer to the mean and variance corresponding to the feature under investigation for the classes C_i, C_j respectively, and m is feature index. As (6.28) tends to emphasizing separation between two classes, we modify this formula to emphasize the feature ranking more than to classes.

$$FDR(m) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K}\left|\mu_i^m - \mu_j^m\right|^2}{\sum_{i=1}^{k} s_i^m} \tag{6.24}$$

In (18) FDR(m) is averaged over all classes. The value of the criterion *FDR(m)* is computed for each of the features. Features are then ranked in order of descending values of *FDR(m)*. The l features corresponding to the l best values of *FDR(m)* are then selected to form the feature vector. The selected top discriminant features are then modeled and recognized by the Hidden Markov Models (HMMs) after some manipulation.

The use of *FDR* is robust because it is independent of the underlying statistical distributions. This approach is similar to the K-L divergence criteria described by Fukunaga [81]. But the divergence generally need constrints on the feature distribution and covariaance matrix for feature selection. When the feature distributions are Gaussian and covariance matrixes are equal, the divergence and FDR criteria are equal.

6.4 Experimental Verifications

6.4.1 Experiment Setup

Experiments were conducted to illustrate the applications of the feature extraction and selection algorithms for micro-milling tool wear state estimation. A total of 27 test experiments are conducted with different spindle speed, depth of cut, radial depth of cut, and feed rate for *Φ500μm* and *Φ800μm* tools (The working conditions are discussed in chapter 8). The materials used are copper and steel.

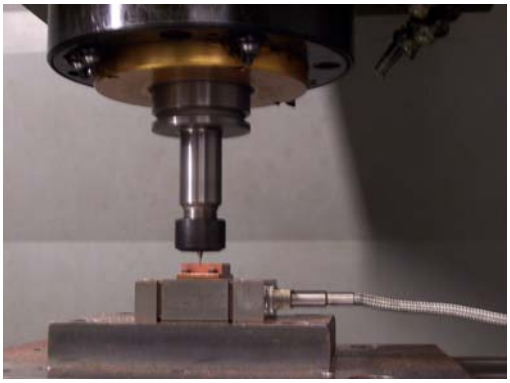


Figure 6.2 Experiment setup

We took the average of the two flute flank wears $VB=1/2(VB_1+VB_2)$ as the final flank wear value, as measured in figure 6.3**.** The progressive change in the flank wear is approximated by a B-spline curve (discussed in section 8.2). Three tool states are defined, and are measured by the threshold that indicates the transition to each state based on the second-order change in the fitted flank wear curve. The average values measured by the thresholds of these three states are listed in table 6.1. We then chose their closest 10μm as the final threshold.

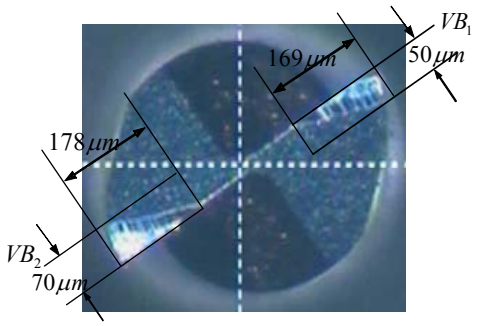


Figure 6.3 Tool wear measurement for Φ500μm

Table 6.1 definition of tool state

Tool diameter	threshold	State 1	State 2	State 3
500 μm	Estimated	17.9μm	17.9-53.8μm	53.8μm
500 μm	Final	0-20 μm	20-50 μm	>50 μm
800 μm	Estimated	23.2μm	23.2-57.5μm	57.5μm
800 μm	Final	0-20 μm	20-60 μm	>60 μm

6.4.2 Feature Normalization and Selection

In this section we illustrate the feature selection performances from principle component analysis (PCA), linear discriminant analysis (LDA) and automatic relevance determination (ARD) with the data from different tool wear states.

For convenience, we denote the coefficient vectors as $d_j(k), j = 1, 2, \ldots J$, where J is the number of wavelet scales, and k represents discrete time. Each coefficient vector can be further represented by its average energy, which is defined as:

$$E_j = \frac{1}{T_j}\sum_{k=1}^{T_j} d_j^2(k) \tag{6.25}$$

where T_j is the number of coefficients at each scale. The value of T_j is different for different scales, because the time spacing of wavelet coefficients varies with the resolution j.

Time domain features are also extracted to show the force statistics. They are the mean, variance, skew, kurtosis and dynamic component and defined in the table 6.2. Please note that the dynamic component one is different from the variance: It measures how far the signal from its medium and it is useful when there are spikes. To be comparable with the features from the wavelet packets, they are rescaled to their $1/32$ times. The sampling rate of the cutting forces is 6,000 Hz. The forces are processed with a segment window of 360 points, amount to 20 revolutions and 0.06 second for 20,000rpm, and 18 revolutions and 0.06 second for 18,000rpm.

Table 6.2 Features extraction and selection scopes

	Feature	Formula	Representation symbol
Time domain	Mean	$\mu = E[f(t)]$	T_1
	Variance	$\sigma^2 = E[(f(t)-\mu)^2]$	T_2
	Skew	$Skew = E[(f(t)-\mu)^3/\sigma^3$	T_3
	Kurtosis	$Kurtosis = E[(f(t)-\mu)^4/\sigma^4 - 3$	T_4
	Dynamic Component	$\Delta f_a(t) = \lvert f_a(t) - f_{med}(t)\rvert$	T_5
Wavelet domain	32 Wavelet Packet Coefficients in the 5-level decomposition	$E_j = \frac{1}{N_j}\sum_{k=1}^{N_j}\left[d_{j,k}^n(t)\right]^2$	$d_{5,1}, d_{5,2}, \ldots d_{5,32}$

Cutting forces are decomposed by the wavelet packet into five levels, and the features are represented by the mean energy of the respective packet node, with $d_{5,1}$ represents first nods of level 5 decomposition. The features from wavelet packets and from time domain, i.e. mean, standard deviation, skew, kurtosis and dynamic component are combined to form a 37-dimension feature vectors.

This high dimension features need to be reduced for less computation and robust representation. Figure 6.4 shows features and the selection result with LDA for the three state tool wear. As shown in the figure, some of the features are quite different (e.g. feature 8, 10 and 21) from the three classes while others are similar (e.g. feature 27, 28, and 29). These reduced dimension features are then input to the Hidden Markov Models (HMMs) for classification.

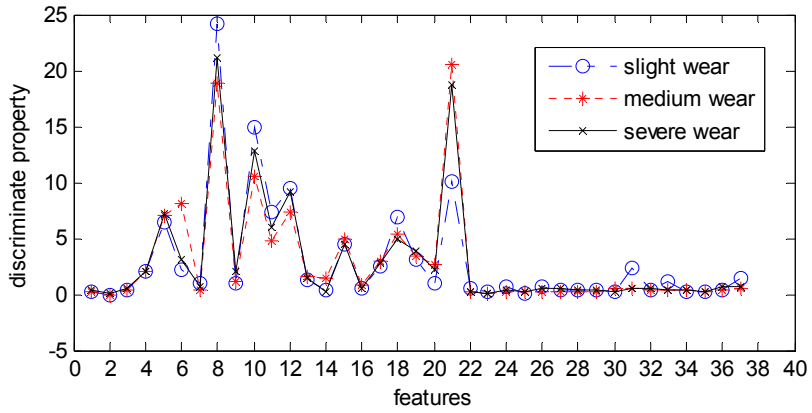
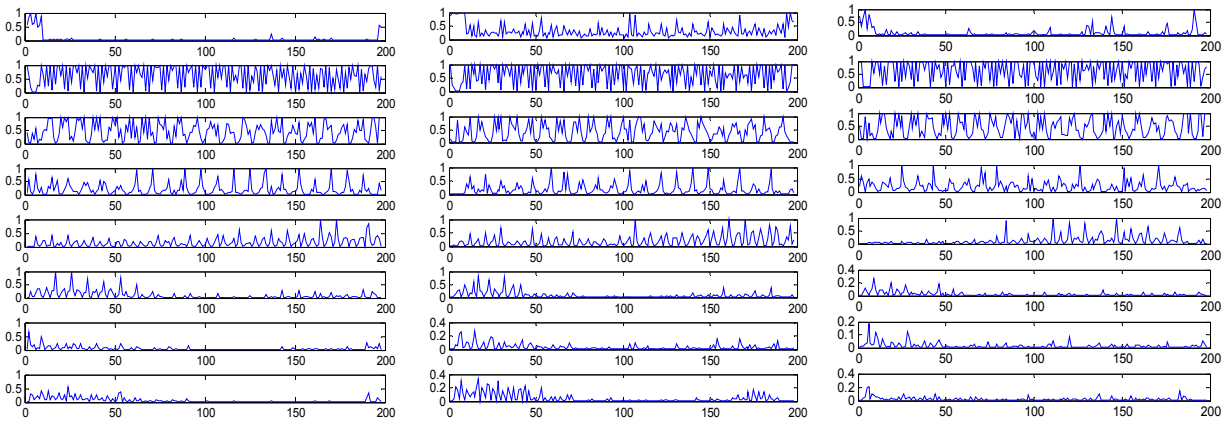


Figure 6.4 Discriminant feature selection

For the real force signal, we are confronted with features vectors that vary within different dynamic ranges because of different cutting conditions. For example, cutting forces are much larger when the depth of cut is higher. Thus, features with large values may have a larger influence in the discriminant function than features with small values, although this does not necessarily reflect their discriminability of the tool states. The problem can be overcome by normalizing the features so that their values lie within similar ranges. The extracted 8-component wavelet feature vectors are normalized between [0, 1]. Figure 6.5 shows the normalized features from different tool wear level. These feature are then will trained for building HMMs or input to HMMs for state estimation.



a) initial wear b) progressive wear c) accelerated wear

Figure 6.5 Normalized top 8 discriminant wavelet packets

6.4.3 HMM for Tool Wear State Classification

For a full discussion on the feature selection, we illustrate the different feature selection approaches with their classification rate with HMMs estimation. The modeling of tool state with HMMs will be discussed in details in Chapter 7. We build an HMM for each tool state for classification, and each HMM is trained with its own feature set. To obtain the HMM for the fresh tool state, for example, we only select the features with tool wear from 0-20μm, and likewise for the subsequent wear states. We then get $\boldsymbol{HMM_1=\lambda_1=(\pi_1, A_1, c_1, \mu_1, \Sigma_1)}$, $\boldsymbol{HMM_2=\lambda_2=(\pi_2, A_2, c_2, \mu_2, \Sigma_2)}$ …, for the corresponding tool wear states. When all the HMMs are trained, we obtain the HMM represented by the parameters: i.e. $\boldsymbol{HMMs=(\lambda_1, \lambda 2\ldots).}$ The Viterbi algorithm is then implemented to estimate the most probable tool wear state sequence. For an observation feature set $\boldsymbol{Y}$, $p_w=p(Y|\lambda_i)$, $i=1,2,\ldots$ is calculated. In the tool state recognition state, the features from an unknown state are extracted and input to the HMMs to match the trained states, and then choose the most probable one.

The HMMs classify the state by maximize a posterior (MAP) probability density, given the features,

$$W_i^* = \arg\max_{w \in W} \{p(W_i = w \mid y_1, \ldots y_i)\} \qquad (6.26\text{ a})$$

or in the physical representation form,

$$p(class\ i \mid features) = \frac{p(features \mid class\ i)\,p(class\ i)}{p(features)} \qquad (6.27\ b)$$

The classification rate is defined as:

$$\frac{Correctly\ classified\ samples}{Misclassifieds\ samples + Correctly\ classified\ samples} \times 100\% \qquad (6.28)$$

6.4.3.1 HMM Classification for Pure Copper

For brief description, we take the representation of

Mean: T_1, variance: T_2, skew: T_3, kurtosis: T_4, dynamic component: T_5

Table 6.3 shows the selection results from the three different approaches with the first 12 highest relevance features from the tests. There are some common features in all the three approaches for example the nods at $d_{5,1}$, $d_{5,5}$, $d_{5,7}$, $d_{5,10}$, $d_{5,13}$ and also some different features such as mean, variance and kurtosis, due to different selection standard. The similarity is measured as the number of common features among all the appeared features.

Table 6.3 Feature selection for copper TCM

Algorithms	Top 12 features with ranking											
	1	2	3	4	5	6	7	8	9	10	11	12
PCA	$d_{5,3}$	$d_{5,16}$	$d_{5,7}$	$d_{5,5}$	$d_{5,13}$	$d_{5,1}$	$d_{5,10}$	$d_{5,12}$	$d_{5,8}$	T_4	$d_{5,21}$	$d_{5,28}$
ARD	$d_{5,1}$	T_1	$d_{5,5}$	T_5	T_2	$d_{5,7}$	$d_{5,8}$	$d_{5,13}$	$d_{5,11}$	$d_{5,10}$	$d_{5,18}$	$d_{5,32}$
LDA	$d_{5,16}$	$d_{5,3}$	$d_{5,5}$	$d_{5,6}$	T_5	$d_{5,7}$	$d_{5,1}$	$d_{5,13}$	$d_{5,10}$	$d_{5,26}$	$d_{5,32}$	$d_{5,29}$
Similarity	Top 4: 10%				Top 8: 30.8%				Top 12: 19.1%			

The classification rate of different feature selection algorithms for pure copper are shown in Table 6.4 and illustrated in figure 6.6. The features in the time domain and wavelet domain combined are selected with LDA, PCA, and ARD. The HMMs classification rare for the top 4-12 best features from different approaches are listed in table 6.4. The different performances are illustrated in figure 8 for comparison. As can be drawn from the figure, the LDA (cross) always has the highest classification rate of all the three approaches and PCA (circle) performs worst. It can also be deduced from the figure that the classification rate increase as the feature dimensions increase at first, and then decrease or oscillate for classification performance. The HMMs classification rates are lowest with only 4-dimension features and also low with 12-dimension features for LDA and PCA, and the ARD performances oscillate with the more feature

dimensions than 8. Another important finding is that all the three approaches have the highest classification rate with 8 feature dimensions, and after that the HMMs perform worse with the increasing of feature dimension. Under the best 8-dimension feature space, the HMMs have the 97.2% classification rate with the LDA(cross) feature selection, while 91.5% and 80.4% with PCA(circle) and ARD(star) selection respectively.

Table 6.4 Classification rate of copper with HMMs

Algorithms	Classification rate with different number of features								
	4	5	6	7	8	9	10	11	12
PCA	0.632	0.693	0.865	0.754	0.804	0.726	0.687	0.636	0.608
ARD	0.561	0.620	0.785	0.850	0.915	0.833	0.894	0.812	0.868
LDA	0.712	0.850	0.875	0.923	0.972	0.962	0.90	0.871	0.864

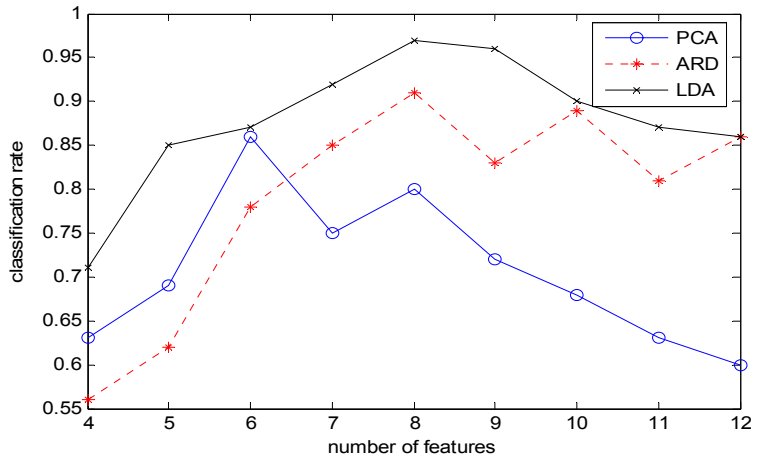


Figure 6.6 Classification rate of different feature selection methods

6.4.3.2 HMM Classification for Steel (T4)

To show the generalization of the feature selection methods, different work piece with steel (T4) working condition for the micro end milling tool wear classification was conducted for verification. Table 6.5 shows the selection results for steel from the three approaches with the first 12 highest relevance features from the tests. There are some common features in all the three approaches for example the nods at $d_{5,1}$, $d_{5,4}$, $d_{5,8}$, $d_{5,10}$, $d_{5,12}$ and also some different features such as mean, variance and kurtosis, due to different selection standard.

Table 6.5 Feature selection for steel TCM

Algorithms	Top 12 features with ranking											
	1	2	3	4	5	6	7	8	9	10	11	12
PCA	$d_{5,2}$	$d_{5,16}$	$d_{5,8}$	$d_{5,4}$	$d_{5,7}$	$d_{5,1}$	$d_{5,10}$	$d_{5,12}$	$d_{5,13}$	T_4	$d_{5,16}$	$d_{5,28}$
ARD	$d_{5,1}$	$d_{5,4}$	T_1	T_5	$d_{5,8}$	T_2	$d_{5,6}$	$d_{5,12}$	$d_{5,11}$	$d_{5,10}$	$d_{5,18}$	$d_{5,32}$
LDA	$d_{5,12}$	$d_{5,2}$	$d_{5,4}$	$d_{5,8}$	T_5	$d_{5,6}$	$d_{5,16}$	$d_{5,1}$	$d_{5,10}$	$d_{5,26}$	$d_{5,32}$	$d_{5,24}$
Similarity	Top 4: 12.5%				Top 8: 33.3%				Top 12: 27.8%			

The top 4-12 features were selected with LDA, PCA, and ARD for the HMMs classification as table 6.5. The classification rare for different approaches are listed in table 6.6 and illustrated in figure 6.7 for comparison. Under this working condition, similar results can be drawn. Of all three feature selection schemes the LDA has the highest classification rate, and PCA performs worst. It can also be deduced from the figure, the classification rate increase as the feature dimensions increase at first, and then oscillate for HMMs classification performance. At the same time, LDA and PCA have the highest classification rates with 8 feature dimensions, and after that the HMMs perform worse with the increasing of feature dimension. Under the top 8-dimension feature space, the HMMs have the 95.8% classification rate with the LDA(cross) feature selection, while 89.5% and 75.4% with PCA(circle) and ARD(star) selection respectively. It has to be mentioned that the ARD has a slightly better classification performance with 9-dimension features 90.3% than with 8-dimension, but the classification rates also decrease after that.

Table 6.6 Classification rate of steel with HMMs

Algorithms	Classification rate with different number of features								
	4	5	6	7	8	9	10	11	12
PCA	0.613	0.685	0.743	0.684	0.754	0.726	0.695	0.626	0.601
ARD	0.581	0.660	0.790	0.889	0.895	0.903	0.894	0.882	0.868
LDA	0.682	0.735	0.834	0.894	0.958	0.916	0.897	0.892	0.864

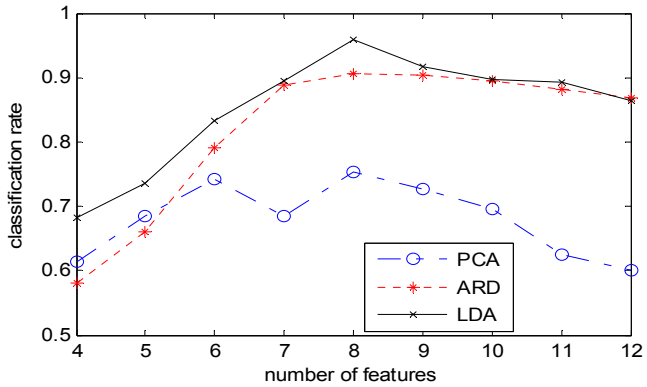


Figure 6.7 Classification rate of different feature selection methods

6.4.4 Discussion

The different performances of PCA, FDA and ARD for feature selection are due to their algorithm properties: PCA is designed for feature dimension reduction, FDA for class discrimination and ARD for correlation of the features to the tool wear states. Unlike those in [181] where ARD performs best and in [130] where PCA has least classification errors, the FDA algorithm suits our tests best because our aim is for state estimation (discrimination), not for feature dimension reduction or for correlation analysis. These inner properties of different algorithms result in different classification rate.

6.4.4.1 ARD vs FDA

Geometric views of ARD versus FDA are figure 6.8(a) and figure 6.8(b). In ARD, we reach a high posterior probability provided data with the prior model, this is Bayesian learning. It only reflects the relation between of input and output, but does not consider the information from other classes. While in FDA, we get the means from two class as far as possible but the within class variance is as close as possible, this is better for classification.

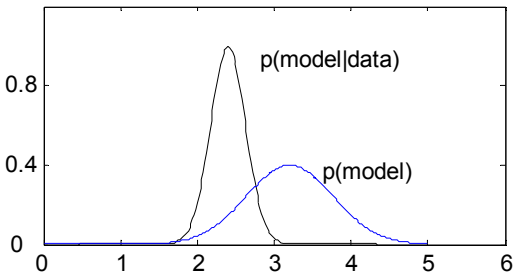


Figure 6.8(a) The Bayesian leaning of ARD: maximizing relation between input and output

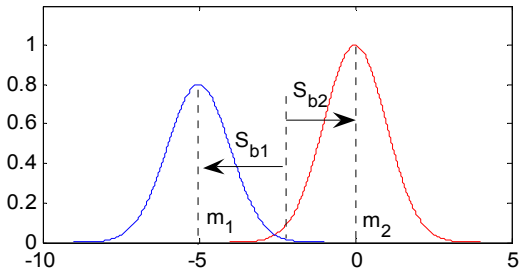


Figure 6.8(b) FDA: Select the class with minimizes within-classes scatter, and maximizes between-classes scatter

Figure 6.8 Comparison between ARD and FDA

6.4.4.2 PCA vs FDA

The differences between PCA and FDA for feature selection are illustrated in figure 6.9(a) and figure 6.9(b). The data are drawn from machining force at different tool wear level. The features are projected according to the algorithms of PCA and FDA. PCA finds the subspace that with highest variance, where only data are considered and no class information needed, it is an unsupervised learning. In PCA, the directions

corresponding to the small eigenvalues are interpreted as “noise", however, it could be in the condition that the information required to perform a classification task lies in the “noise” dimensions thrown away. FDA finds the directions that maximize the separability. It includes the class information and is a supervised learning.

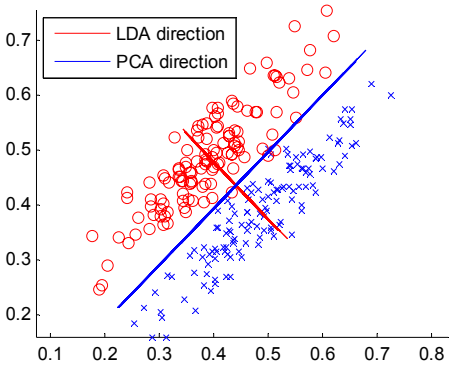


Figure 6.9(a) shows force data and FDA (red circle) and PCA (blue cross) directions

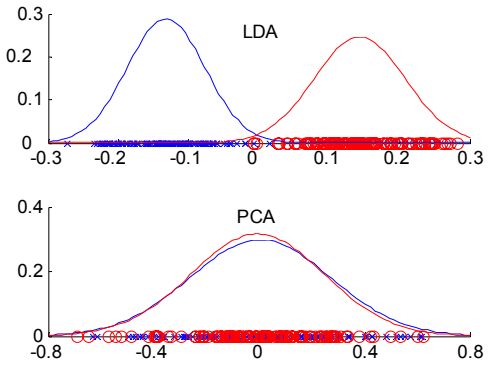


Figure 6.9(b) shows the projected force data in FDA and PCA directions

Figure 6.9 Comparison between PCA and FDA

6.5. Conclusion

In this chapter, Fisher's linear discriminant analysis is adapted for feature selection in HMM modeling. The FDR method is robust because there are no assumptions made with regard to the features' distributions. Features from both time and wavelet domain are ranked only on their class-separation ability. Selecting the most discriminant features according to their FDR scores improve HMMs for TCM. The effectiveness of the approach was evaluated under different working conditions. Successful tool state detection was compared with other feature selection methods such as ARD and PCA. Experimental results indicate that the selection of top 8 most discriminant features can achieve the highest classification, average classification rate (for 3-state tools) as high as 97.2% can be achieved. Therefore, the FDR approach is a successful feature selection tool for HMM modeling in TCM.

Chapter 7
Continuous Hidden Markov Models for Micro-milling Tool Wear Classification

In this chapter, a modeling framework based on hidden Markov models (HMMs) is implemented to classify force features of the different tool states. Previous application of HMMs for drilling [20] [54] [55][70] and turning [171] [203] are not suitable to our problems. For HMMs in milling, the important issues for modeling HMMs, such as the selection of the number of Gaussian mixture components and HMM states, were not discussed in details in [14] [15] [29][56][89][149]. The HMM implemented in this study is based on the concept of multi-rate modeling of [56], but with several modifications to adapt to the conditions in micro-milling monitoring, which is discussed in the later sections accordingly.

The chapter is organized as follows: Section 7.1 introduces the formal definition of Hidden Markov Models and introduces the three problems of Hidden Markov models and their solution. Section 7.2 presents a framework for carrying out tool wear classification using Continuous HMMs. The selection and modeling issues of HMMs are also discussed with details in this section. Section 7.3 shows the HMM estimation results from micro-milling experiments, and section 7.4 offers a discussion regarding the limitations of HMMs and their generalization potential.

7.1 Hidden Markov Models

The Markov process is a stochastic process that is used to model the evolution of random state as function of time. For the first-order Markov process for a system with *S* distinct states, the current state of the model is dependent only on previous state,

$$P\left(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots\right) = P\left(q_t = S_j \mid q_{t-1} = S_i\right) = a_{ij} \qquad (7.1)$$

where q_t is actual state at time t and a_{ij} is the transition probability between state i and j. The transition constraints to,

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{7.2}$$

The hidden Markov model (HMM) is a double-layered stochastic process, with an underlying finite-state hidden Markov process that associates with the observation process. For our application, the force features $Y = \{y_1, y_2, ..., y_T\}$ are the observation process, and the tool state sequence $S = \{s_1, s_2, ..., s_T\}$ is the hidden process. The probability of the observations Y is conditionally dependent on state S,

$$P(Y) = \sum_{S} P(Y, S) = P(y_1|s_1)P(s_1)\prod_{t=2}^{T} P(y_t|s_t)P(s_t|s_{t-1}) \tag{7.3}$$

In order to characterize an HMM completely, the following elements are needed. An HMM is identified with the parameter set $\lambda = (\pi, A, B)$ where

π is the initial state distribution vector, e.g. π is the probability of state i at the same arbitrary time $t = 0$.

A is the state transition matrix, where $A = [a_{ij}]$, with a_{ij} being the probability of transiting to state j given current state i.

B is the output distribution matrix, where $B = [b_{jk}]$, with b_{jk} being the probability of observing feature k given current state j.

The estimation and selection of parameter set $\lambda = (\pi, A, B)$ are the essence of HMM modeling, which are discussed in the following sections.

In HMM based TCM, it is assumed that the sequence of features corresponding to each tool state is generated by a Markov model as shown in Fig. 7.1. A Markov model is a finite state machine which changes state once every time unit and each time t that a state j is entered, a feature vector o_t is generated from the probability density $b_j(o_t)$. Furthermore, the transition from state i to state j is also probabilistic and is governed by the discrete probability a_{ij}. Fig. 7.1 shows an example of this process where the six

state model moves through the state sequence Q = 1*;* 2*;* 2*;* 3*;* 4*;* 4*;* 5*;* 6 in order to generate the sequence o_1 to o_6.

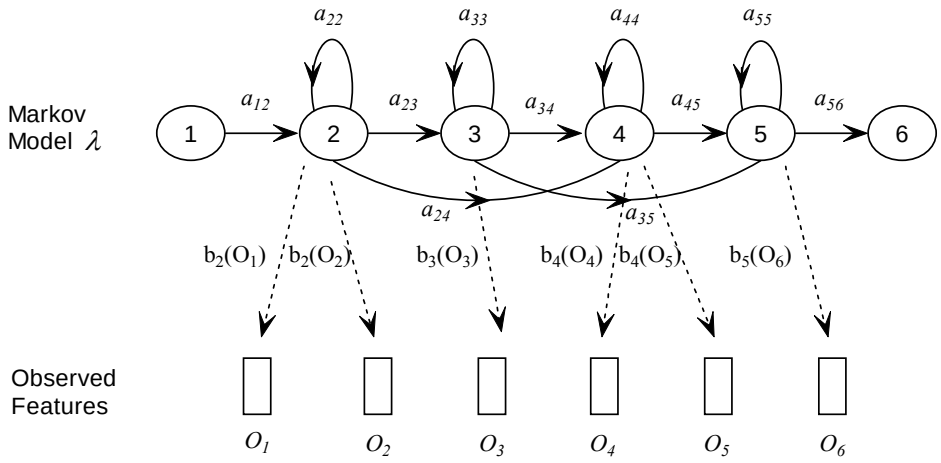


Fig. 7.1 the HMM Generation Model

7.2 Three Problems of Hidden Markov Models

Given the definition of HMMs above, three problems of interest must be addressed before they can be applied to real-world applications. They listed below:

1. The Evaluation Problem: Given a model λ and a sequence of observations $O = (o_1, o_2, ..., o_T)$, what is the probability $P(O \mid \lambda)$ namely, the probability of the model that generates the observation?

2. The Decoding Problem: Given a model λ and a sequence of observations $O = (o_1, o_2, ..., o_T)$, what is the most likely state sequence $S = (q_0, q_2, ..., q_T)$, in the model that produces the observations?

3. The Learning Problem: Given a model λ and a set of observations, how can we adjust the model parameter λ to maximize the joint probability (likelihood) $\prod_o P(O \mid \lambda)$.

If we could solve the evaluation problem, we would have a way of evaluating how well a given HMM matches a given observation sequence. Therefore, we could use

HMM to do pattern recognition, since the likelihood $P(O/\lambda)$ can be used to compute posterior probability $P(\lambda/O)$, and the HMM with highest posterior probability can be determined as the desired pattern for the observation sequence. If we could solve the decoding problem, we could find the best matching state given an observation sequence, in other words, uncover the hidden state sequence. If we could solve the learning problem, we would have the means to automatically estimate the model parameter λ from an ensemble of training data. These three problems are tightly linked under the same probability framework. The efficient implementation of these algorithms share the same principle of dynamic programming. The detail description of the solution to these problems are attached in appendix C.

7.3 Hidden Markov Models Based Tool Condition Monitoring

7.3.1 HMM Description of Tool Wear Process and Monitoring

Like conventional milling, tool wear increases progressively with machining time, given specific working conditions in micro-milling. Generally the dynamic process of the wear rate $v(t)$ can be modeled as a differential equation:

$$v(t) = \frac{dw(t)}{dt}, \tag{7.4}$$

where $w(t)$ is the tool wear value at machining time t.

Approximating this function with difference function, we get

$$\frac{w(t+\Delta t) - w(t)}{\Delta t} = v(t) \quad \Rightarrow \quad w(t+\Delta t) = w(t) + v(t)\Delta t \tag{7.5}$$

For convenience and to facilitate physical interpretation, we rescale Δt to unit time step 1 and keep $v(t)$ as constant v but including a noise term $\varepsilon(t)$, which obeys a certain distribution.

$$w(t+1) = w(t) + v + \varepsilon(t) \tag{7.6}$$

If $\varepsilon(t)$ obeys a Normal distribution $N(\mu, \sigma^2)$ with mean μ and variance σ^2, then

$$p(w(t+1) \mid w(t)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(w(t+1) - w(t) - v)^2] \quad (7.7)$$

This means that we can predict the tool wear value $w(t+1)$ at time $t+1$ given the current value $w(t)$ and wear rate. This is a first-order Markov process which refers that the current state only depends on the immediately preceding state in time. More formally,

$$p(w(t+1) \mid w(t), w(t-1), ...w(1)) = p(w(t+1) \mid w(t)) \text{ (hidden process) } (7.8)$$

This process is hidden to us, and hence classed as *hidden process*, which we try to uncover. If the wear rate depends on second-order differential equations under the current wear, we can likewise approximate it with a second-order Markov process description, and this can be similarly repeated for higher-order representation. The second and higher-order Markov Model can describe more dynamics of the wear process but result in more complex models and more parameters to estimate. In this work, the first-order Markov Model is used to model the dynamic system.

For online estimation, we cannot measure the current tool wear value directly, but estimate it from the observation of the force features $y(t)$. This relationship is very complex, and typically non-stationary and uncertain. It is a function of $w(t)$:

$$y(t) = f(w(t)) \qquad \text{(observation process)} \qquad (7.9)$$

which represents the relationship of a certain feature $y(t)$ and a tool wear value $w(t)$. $y(t)$ is generally extracted from measurable sensor signals, so we call it the *observation process*.

The above double-layer stochastic processes can be illustrated using figure 7.2. The top three figures are the observed force features corresponding to different tool states shown in the pictures below. The variations are stochastic, while the upper is non-stationary and the bottom is stationary and Markov. Their relationships are modeled with HMMs, which will be discussed in details in later sections.

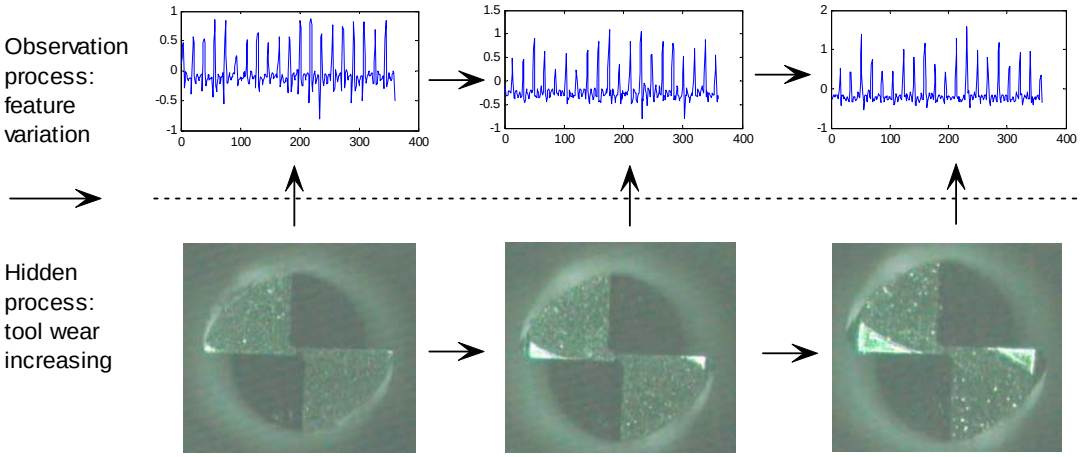


Figure 7.2 Stochastic modeling of tool wear process

We can formally specify the TCM problem as trying to find the most probable state given the machining signal features.

The aim of TCM is to find,

$$\text{TCM: } \arg\max_{tool\ state\ i} p(tool\ state \mid signal\ features) \tag{7.10}$$

This is a dynamic inference problem since we do not estimate state only with prior knowledge, but also adapt to the current features. In the framework of HMMs, this kind of dynamic reasoning problems can be solved.

7.3.2 The Framework of HMMs for TCM

The framework of HMMs for TCM is illustrated in figure 7.3. It first decomposes the signal into different wavelet packets, and selects the discriminant wavelet packets, and then uses these packets as features for HMMs training. Once the models are trained, we get the $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$ that represent different tool wear states. In the tool state recognition stage, the extracted features from an unknown state are input to the HMMs to match the trained states, and then choose the most probable one.

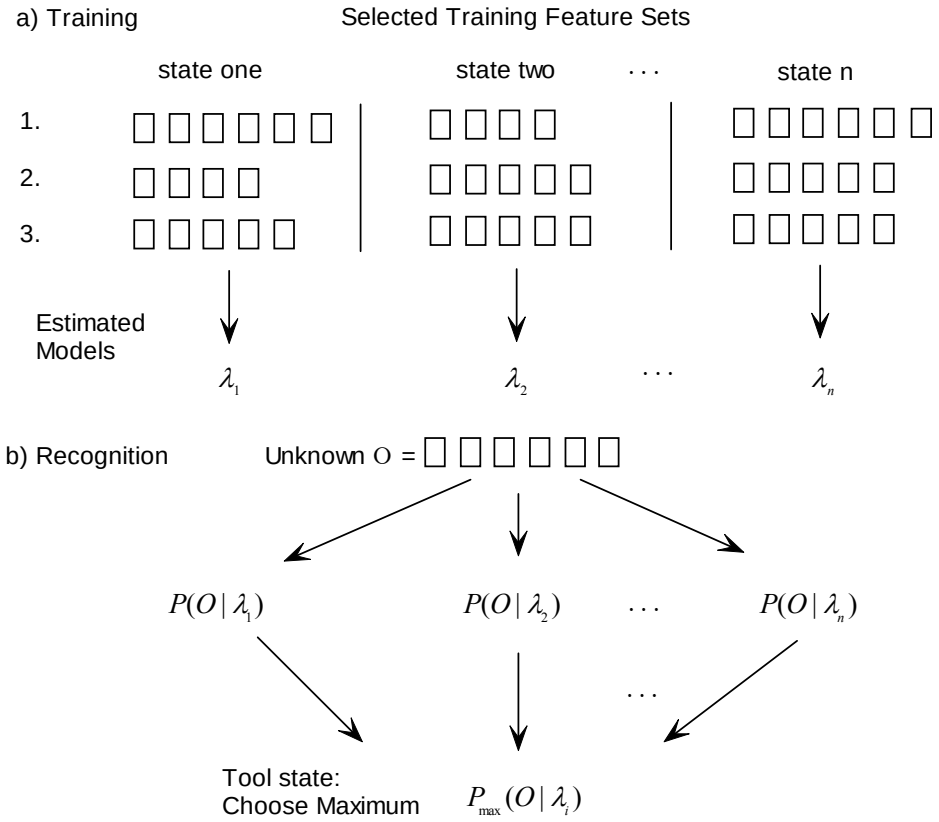


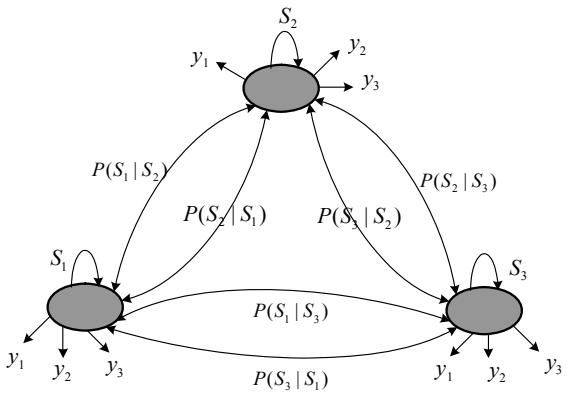
Figure 7.3 Framework of Hidden Markov Models for tool wear classification

7.3.3 Hidden Markov Model Selection: Continuous Left-Right HMMs

7.3.3.1 Left-Right HMMs

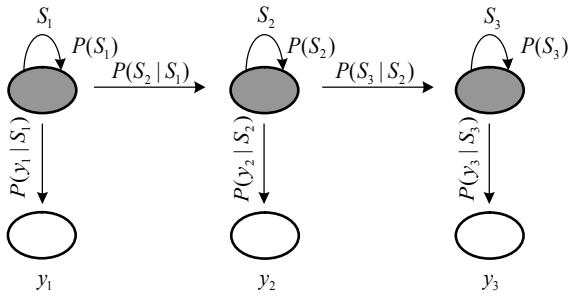
The most widely used HMMs are the ergodic and the left-right HMMs. The ergodic HMM is a generic type of HMM and adapt to many kinds of problems. The ergodic HMM is fully connected with all states linked together, which infers that every state can be reached from every other state. The transition matrix is a full matrix as shown in figure 7.4).

However, ergodic HMMs are not feasible for TCM because the machining process is continuously changing and the state can not change back. Hence, there is the need to constrain the transition matrix (as shown in figure 7.5), the left-right HMM can transit to self and later state, but not to the former state. This transition matrix is thus constrained to an upper triangular matrix. The initial state of left-right HMM is $\pi_0 = [1 \;\; 0 \;\; 0]$ since the tool starts at the first state.



Transition matrix: $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} a_{ij} \in [0,1]$

Figure 7.4 ergodic HMM tool state model



Transition matrix:

$$\begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}, \quad a_{ij} = P_{ij} = 0 \ \ for \ \ |i-j| > 1 \ \& \ i < j$$

Initialization: $\pi_0 = [1 \ \ 0 \ \ 0]$

Figure 7.5 left-right HMM tool state model

Not like that of [41], the HMMs involves no jumping between states but progresses from one to another, i.e. the tool starts in the initial wear state, moves into the gradual wear state, which is relatively long in micro-milling, before advancing into the accelerated wear state.

7.3.3.2 Continuous HMMs and Gaussian Mixtures Modeling

Besides classified with structures, the HMMs also classified discrete HMMs or continuous HMMs depending on whether the observation features are modeled as

continuous or discrete. In the case of a discrete HMM, the feature space ***Y*** consists of a finite number of elements, $\boldsymbol{e_1}$,.....,$\boldsymbol{e_M}$, which are achieved through the Vector Quantization (VQ) [160]. For most applications, VQ of the features can degrade performance significantly since it is an average over large regions. Moreover, the codebooks generated by the quantization process are constructed using training data from all classes. When a new class of shapes is added, we need to reconstruct the codebook and retrain all system modules. We choose continuous HMMs for our study.

In a continuous HMM, the feature (observation) space ***Y*** is considered to be infinite and continuous. Therefore, for each state ***j*** we aim to compute the probability $\boldsymbol{b_j(y)}$ for each vector ***y*** in the space. For a probability distribution mentioned above, however, it is not sufficiently flexible to accurately model the variation which occurs between different feature vectors which correspond to a state. This is particularly true if the models are used to characterize wear state from a number of features. Thus, Gaussian mixtures are typically used to model broad sources of variability. An ***K***-component Gaussian mixture is a linear combination of ***K*** Gaussian densities.

The *d*-dimensional Gaussian probability density function (pdf):

$$N_{(\mu,\Sigma)}(x) = \frac{1}{\sqrt{2\pi}^{d}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^{T}(x-\mu)} \tag{7.11}$$

with mean vector μ and covariance matrix Σ.

The most general representation of the PDF is a finite mixture of normal distributions with different means and variances for each state.

$$b_j(y) = \sum_{m=1}^{M} C_{jm} N(y, \mu_{jm}, \Sigma_{jm}) \quad 1 \le j \le N \tag{7.14}$$

where C_{jm} is mixture weights for the m-th mixture in state *j* and the weights are all positive and sum to one. N is Gaussian with mean μ_{jm} and covariance matrix Σ_{jm} for the m-th mixture component in state *j* .

By varying the number of Gaussians *K*, the weights c_k , and the parameters μ_k, Σ_k of each Gaussian density function, Gaussian mixtures can be used to describe any

complex density function. Figure 7.6 is an example of the wavelet feature vector modeled as Gaussian Mixture. Figure 7.6 a) is the original one dimensional signal and approximated with three weighted Gaussian components in figure 7.6 b).

$$f(x) = 0.3 \times N_1(\mu_1, \Sigma_1) + 0.2 \times N_2(\mu_2, \Sigma_2) + 0.5 \times N_3(\mu_3, \Sigma_3)$$

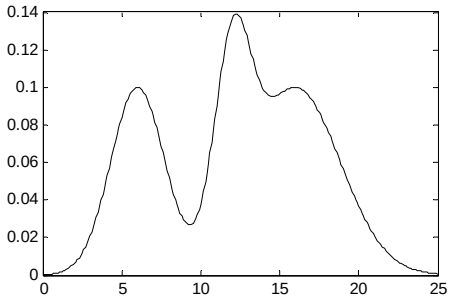


Figure 7.6 a) original signal

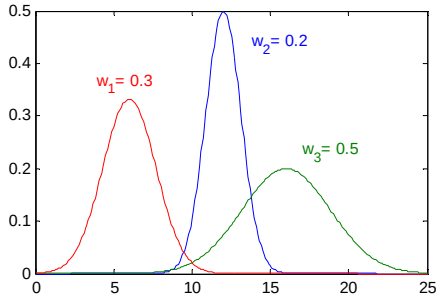


Figure 7.6 b) approximated with 3 Weighted Gaussian Mixtures

7.3.4 Selection of the Number of Gaussian Mixture Components

To model the wavelet feature vectors as Gaussian mixtures, the number of mixture component has to be decided. Theoretically a GMM can approximate any signal to a certain precision when provided with enough mixture components. Too many components may not be desirable, however, as they over fit the data and involve needless computation. There are many criteria for the selection of the components, including Akaike information criterion (AIC), Bayesian information criteria (BIC), and Cross Validation (CV). The Bayesian Information Criterion (BIC) provides a reliable approximation to the integrated likelihood and most popular among all the criteria [57]. We apply the BIC for the selection. The Bayesian information criterion imposes a penalty for including too many terms in a regression model.

The Gaussian mixture model is characterized by the number of components K, the component weights w, and the vector mean μ and covariance Σ: $\theta = (w_1, w_2, ...w_K, \mu_1, \mu_2, ...\mu_K, \Sigma_1, \Sigma_2, ...\Sigma_K)$. The way for choosing a model is to select this one maximizing the likelihood,

$$\hat{K} = \arg\max\, p(y \mid K) \qquad (7.15)$$

An asymptotic approximation of the integrated likelihood leads to minimize the so called BIC criterion

$$BIC_K = -2L_K + v_K \ln n \tag{7.16}$$

where v_K is the number of free parameters with K components, n is the number of observations, $L_K = \log f(y \mid K, \hat{\theta})$ is the maximum log-likelihood for K, and $\hat{\theta}$ is the ML estimate of θ.

BIC selects the best model for the data by balancing the complexity of the model with its fit to the data. The preferred model is the one with the highest value of the criterion. The larger the BIC value, the better the model. The number of Gaussian components is identified to be 9 among the training tests for the 8-dimension feature vectors. For 9 components, the BIC information reaches the highest as illustrated in the figure 7.7.

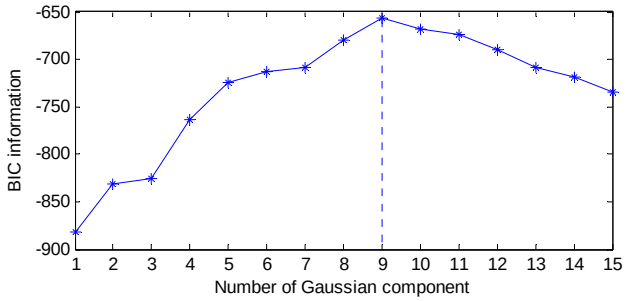


Figure 7.7 Bayesian information criterion for mixture components

7.3.5 On the Number of Hidden States in Each HMM

The selection of appropriate number of hidden states in the HMM is from a physical point of view. Generally with one HMM modeled for one tool state, the states of HMMs lack of the physical meaning as a single HMM for different tool wear states. The tool wear state is characterized by the corresponding entire HMM in our models. The states in that HMM are the Markov transition state within the HMM, but not the tool wear states.

In the milling, the feature patterns are quite different even in the same tool wear state but among different cutting stage. So we train HMMs with the data based on each

cutting pass, and constrain every HMM with three states as: entry, progressing, and exit. In this sense, the HMMs can estimate both the tool wear state (i.e., which HMM match most) and the cutting state (i.e., which state in that HMM highest probable).

7.3.6 Estimation of the HMM Parameters for Tool Wear Classification

For the Gaussian mixture HMMs, the HMMs are characterized with the parameter set $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Once the number and topology of HMM states is chosen as discussed above, the distribution parameters are to be trained.

The Training problem can be described as: Given the observed features $\{y_i\}$, determine the HMM parameters λ that best characterize the features. These parameters can be iteratively estimated by the Expectation Maximization (EM) algorithm (Baum-Welch algorithm) [160].

In the Continuous HMM, the EM formulae find a new set of parameters in terms of the variables calculated by the Forward-Backward procedure [160]. The forward variable is defined as the probability of the partial observation sequence $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_t$ (until time t) and state S_i at time t, given the model λ, as $\alpha_t(i)$. And the backward variable is defined as the probability of the partial observation sequence from $t+1$ to the end, given state S_i at time t and the model λ, as $\beta_t(i)$. Both $\alpha_t(i)$ and $\beta_t(i)$ are determined using the following forward-backward procedure:

Forward:

$$\alpha_1(i) = \pi_1 * b_i(Y_1) \tag{7.17}$$

$$\alpha_{t+1}(j) = \sum_i [\alpha_t(i) * a_{ij}] b_j(Y_{t+1}) \tag{7.18}$$

Backward:

$$\beta_T(i) = a_{iN} \tag{7.19}$$

$$\beta_{t-1}(i) = [\sum_j \beta_t(j) * a_{ji}] b_i(Y_{t-1}) \tag{7.20}$$

The probability of the observations sequence is:

$$P(Y \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) = \sum_{i=1}^{N} \alpha_T(i) \tag{7.21}$$

And the probability of being in state S_i at time t, given the observation sequence $\boldsymbol{Y}$, and the model λ, as:

$$\gamma_t(i,k) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \cdot \frac{c_{jm} N(Y_t, \mu_{jm}, \Sigma_{jm})}{\sum_{m=1}^{M} c_{jm} N(Y_t, \mu_{jm}, \Sigma_{jm})} \tag{7.22}$$

The probability of being at state S_i at time t and state S_j at time $t+1$ is:

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(Y_{t+1}) \beta_{t+1}(i)}{P(Y \mid \lambda)} \tag{7.23}$$

Each of the model parameters $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is maximized separately with EM, by setting its respective partial derivative equal to zero. The results are:

$$\pi_i = \gamma_0^i = [1\ 0\ 0] \tag{7.24}$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T} \gamma_t(i,j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \gamma_t(i,j)} \tag{7.25}$$

$$\overline{c}_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T} \sum_{j=1}^{M} \gamma_t(i,j)} \tag{7.26}$$

$$\overline{\mu}_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j) Y_t}{\sum_{t=1}^{T} \gamma_t(i,j)} \tag{7.27}$$

$$\overline{\Sigma}_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j) \cdot [(Y_t - \mu_{ij})(Y_t - \mu_{ij})']}{\sum_{t=1}^{T} \gamma_t(i,j)} \tag{7.28}$$

It is noted that Eq. 7.25 is essentially the ratio between the expected number of transition from state i. For the output probability re-estimation Eq. 7.26-27.8, the numerator is the expected number of times the observation data emitted from state j

with the observation symbol o_k, and the denominator is the expected number of times the observation data emitted from state *j*.

According to the EM algorithm, the Baum-Welch (Forward-backward) algorithm guarantees a monotonic likelihood improvement on each iteration and eventually the likelihood converges to a local maximum.

Table 7.1 Expectation-Maximization

1. Estimate an initial HMM model as $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.
2. Given λ and the observation sequence $\boldsymbol{Y}$, we calculate a new model $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $P(Y \mid \bar{\lambda}) > P(Y \mid \lambda)$.
3. If the improvement $\frac{P(Y \mid \bar{\lambda}) - P(Y \mid \lambda)}{P(Y \mid \bar{\lambda})} < threshold$, *then stop*

   Put $\bar{\lambda}$ instead of λ and go to step 1.

7.3.7 Tool State Estimation with HMMs

The forward algorithm computes the probability that an HMM generates an observation sequence by summing up the probabilities of all possible paths, and it does not provide the best sate sequence. It is desirable to find such a path. Since the state sequence is hidden in the HMM framework, the most widely used criterion is to find the state sequence that has highest probability of being taken while generating the observation features. In other words we are looking for the state sequence $S = (s_1, s_2, ...s_T)$ that minimizes $p(S, Y \mid \lambda)$. A formal technique based on dynamic programming, know as Viterbi algorithm can be used to find the best path sequence for an HMM.

The Viterbi algorithm can be regarded as the dynamic programming algorithm applied to the HMM or as a modified forwarding algorithm. Instead summing up probabilities from different paths coming to the same destination state, the viterbi algorithm picks and remembers the best path. The best score accounts for the first *t* observations and ends in state S_i, which can be expressed as follows:

$$\delta_t(i) = \max P(q_1 q_2 ... q_t = i, Y_1 Y_2 ... Y_t \mid \lambda) \tag{7.47}$$

The complete procedure for finding the best state sequences is as follows. (ψ is the variable that track the argument which maximized)

.

Table 7.2 Viterbi Algorithm

Initialization:

$$\delta_1(i) = \pi_i b_i(Y_1) \qquad 1 \le i \le N$$
$$\psi_1(i) = 0$$

Recursion

$$\delta_t(i) = \max\left[\delta_{t-1}(i) a_{ij}\right] b_j(Y_t) \quad 2 \le t \le T$$
$$\psi_t(i) = \arg\max\left[\delta_{t-1}(i) a_{ij}\right] \quad 1 \le i \le N$$

Termination

$$P^* = \max\left[\delta_T(i)\right] \quad 2 \le t \le T$$
$$q_T{}^* = \arg\max\left[\delta_T\right] \quad 1 \le i \le N$$

Path backtracking

$$q_T{}^* = \psi_{t+1}(q_{t+1}{}^*) \quad t = T-1, T-2, ...1$$

As shown in Fig. 7.8, this algorithm can be visualized as finding the best path through a matrix where the vertical dimension represents the states of the HMM and the horizontal dimension represents the frames of speech (i.e. time). Each large dot in the picture represents the log probability of observing that frame at that time and each arc between dots corresponds to a log transition probability. The log probability of any path is computed simply by summing the log transition probabilities and the log output probabilities along that path. The paths are grown from left-to-right column-by-column.

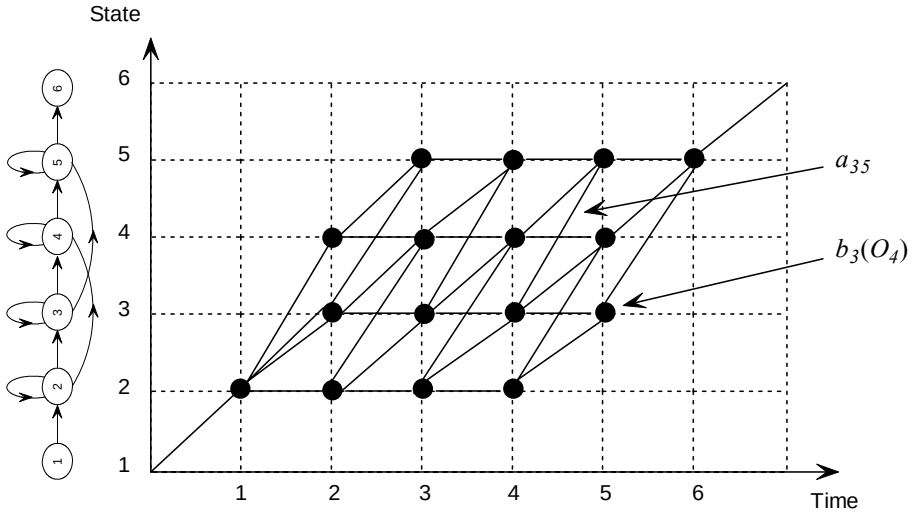


Figure 7.8 The Viterbi Algorithm for State Recognition

7.4 Experimental Verifications

7.4.1 Experiment Setup

The experiment setup is shown in figure 7.9. The work piece was clamped on the dynamometer on the feed table of the machining center with a holder. The cutting force was measured with Kistler 9256A dynamometer. Tool wear is measured using the Olympus Toolmakers microscope, 213 times enlargement. The cutting force output was recorded on a Sony digital tape recorder. Cutting conditions: Feed-rate: 50-180 mm/min; spindle speed: 20,000/18,000 rpm/min; depth of cut: 30–300 μm. Cutter: C-CHES 2005-0150, diameter Φ500μm; C-CHES 2005-0180, diameter Φ800μm. (The working conditions are discussed in section 8.1).

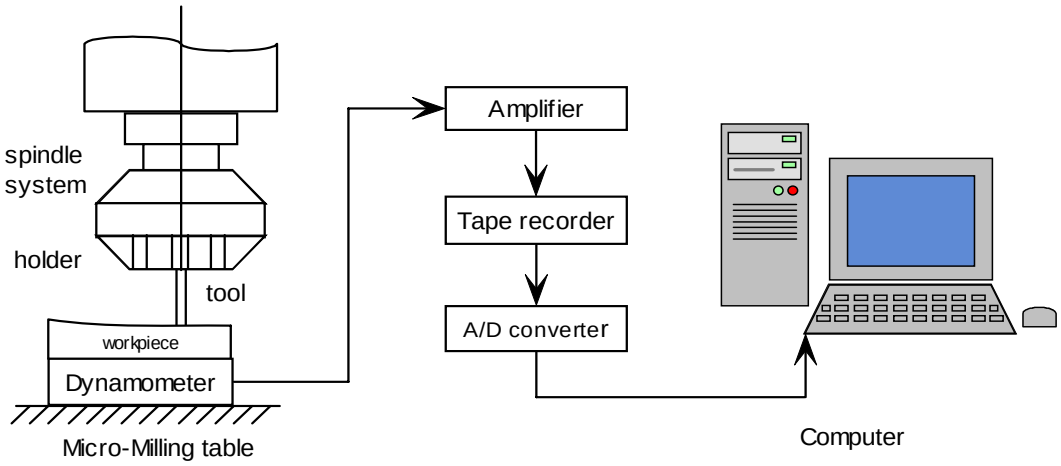


Table 7.9 Experiment setup

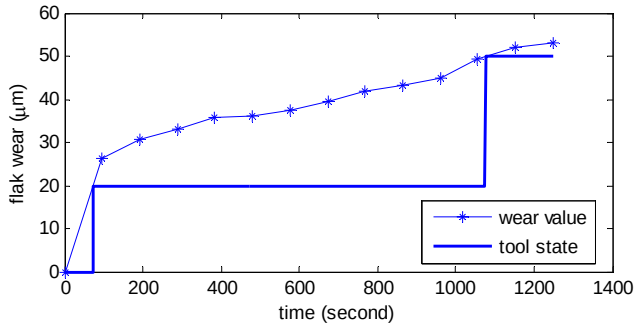
A total of 27 test experiments are conducted with different spindle speed, depth of cut, radial depth of cut, and feed rate for *Φ500μm* and *Φ800μm* tools. The materials used are either copper or steel.

7.4.2 HMM training for TCM

As mentioned above, we train each HMM with its own training sets. To train the HMM that represents the fresh tool state, we only use the feature vectors from that state, and similarly for the medium wear state and severe wear state.

We apply the EM algorithm to estimate the HMM parameter $\boldsymbol{\lambda}=(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The algorithm is implemented to interactively adjust the model parameter to maximize the probability of the observation sequence. The EM method works as discussed in section 7.3.

Figure 7.10 shows the training set for the HMM modeling, whose working conditions are listed in the right side. The selection of the training set is based on the distance to the mean working parameters. The EM algorithm converges quickly as illustrated in figure 7.10 from test 4.



Test 4, Copper
Spindle speed: 20,000 rpm
DOC: 60μm
Radial DOC: 75μm
Feed rate: 120mm/min

Figure 7.10 a) Tool wear state for training the HMMs

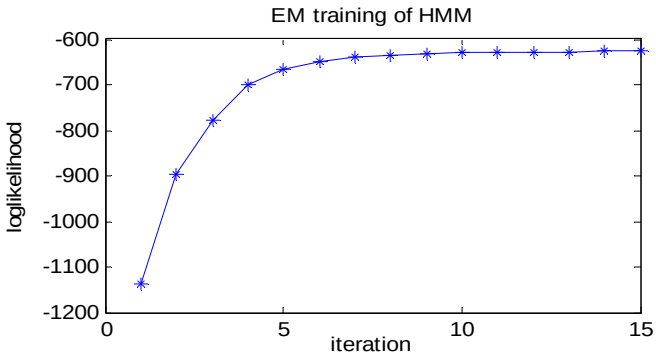


Figure 7.10 b) Convergence of EM for HMMs training

Figure 7.10 HMM training for test 4

Other working conditions are also trained with to build HMMs in the same way. They are the test 10, test 13, test 20 and test 24, which will be discussed in chapter 8.

7.4.3 HMM for tool wear state estimation

Because there are 5 HMMs trained, we need to decide which HMMs suit best for the tool state estimation. Given the working condition, a similarity measure based on the Euclidian distance of the working condition is first determined as follows:

$$d_i = \left[\left(\frac{v - v_{ri}}{v_{ri}}\right)^2 + \left(\frac{f - f_{ri}}{f_{ri}}\right)^2 + \left(\frac{d_v - d_{vri}}{d_{vri}}\right)^2 + \left(\frac{d_r - d_{rri}}{d_{rri}}\right)^2 \right]^{1/2} \qquad (7.45)$$

where d_i is the distance, v_{ri} is the *i-th* reference spindle speed, f_{ri} is the *i-th* reference feed rate, d_{ri} is the *i-th* reference vertical depth of cut, and d_{rri} is the *i-th* reference radial depth of cut. Our aim is to find the one with the minimum distance, $\min d_i$. The other two variables, i.e. work piece materials and tool diameters, are not included here as they are usually predetermined and assumed to be the same.

Wavelet features from various time-scale distributions are extracted and input to the trained HMM for recognition. We build an HMM for each tool state for classification, and each HMM is trained with its own feature set. To obtain the HMM for the fresh tool state, for example, we only select the features with tool wear from 0-20μm, and likewise for the subsequent wear states. We then get $\boldsymbol{HMM_1=\lambda_1=(\pi_1, A_1, c_1, \mu_1, \Sigma_1)}$, $\boldsymbol{HMM_2=\lambda_2=(\pi_2, A_2, c_2, \mu_2, \Sigma_2)}$ …, for the corresponding tool wear states. When all the HMMs are trained, we obtain the HMM represented by the parameters: i.e. $\boldsymbol{HMMs=(\lambda_1, \lambda 2\ldots)}$. The Viterbi algorithm is then implemented to estimate the most probable tool wear state sequence. For an observation feature set $\boldsymbol{Y}$, $p_w = p(Y \mid \lambda_i),\ i = 1,2,\ldots$ is calculated.

The unknown state is then uncovered with the model given the maximum probability of the observed features:

$$w^* = \arg\max\ p_w, w = 1,2,\ldots \tag{7.46}$$

Thus, w^* is the optimum class for the observation $\boldsymbol{Y}$.

The tool state estimation problem can be formalized as: *find* $\boldsymbol{\lambda}$ *= argmax P(Y|* $\boldsymbol{\lambda}$*).* Given a fixed HMM with parameters, determine the most likely sequence of hidden states $\{s_i\}$ for an observed set of features $\{y_i\}$. This decision process is made by *Viterbi Algorithm.* Analogous to the use of HMMs in speech recognition, the classification of the wear level consists of finding the best alignment of the feature vectors to the HMM states via the Viterbi algorithm, which finds the most likely state sequence using dynamic programming. For each segment, a decision was made on the wear state based on the full history of observations $Y = \{Y_1, Y_2, \ldots Y_i\}$. The classification $\hat{W}_i$ corresponds to the most likely state sequence.

The best score accounts for the first *t* observations and ends in state S_i, which can be expressed as follows:

$$\delta_t(i) = \max P(q_1 q_2 \ldots q_t = i, Y_1 Y_2 \ldots Y_t \mid \lambda) \tag{7.47}$$

7.4.4 Moving Average for Tool Wear State Estimation Smoothing

In the multi-rate modeling, the classification the tool states are also constrained to the left-right direction, that is, tool states change from initial to progressive and then to accelerated region but can not change back, the tool state may pre-enter into the next state because of the high noise level in the signal in micro machining. We eliminate this situation by introducing a 15-point moving medium filter. In the HMM recognition, we hold the consecutive 15-state estimations but not decide the state immediately, and then find the medium of these 15 estimates. This medium state is our final tool wear state. Figure 7.11 illustrates the improvement of this approach. It is observed that low classification rate due to misclassifications of the neighboring states in 7.11 a) and the early state changing of the HMMs in figure 7.11 b). After the left-right constraint and medium filter (figure 7.11 c)), the classification improves from 89.6% to 96.5% for test 12.

The classification rate is defined as:

$$\frac{Correctly\ classified\ samples}{Misclassifieds\ samples + Correctly\ classified\ samples} \times 100\% \qquad (7.48)$$

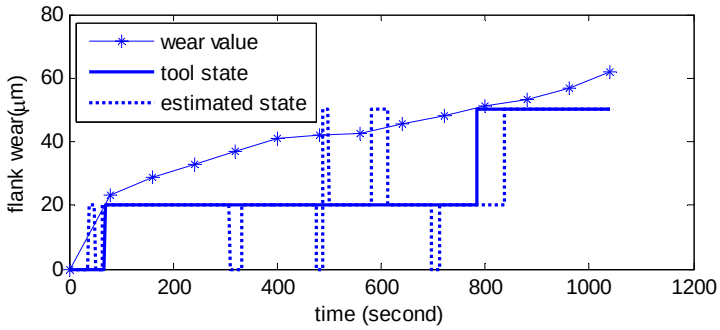


Figure 7.11 a) The general HMMs for TCM

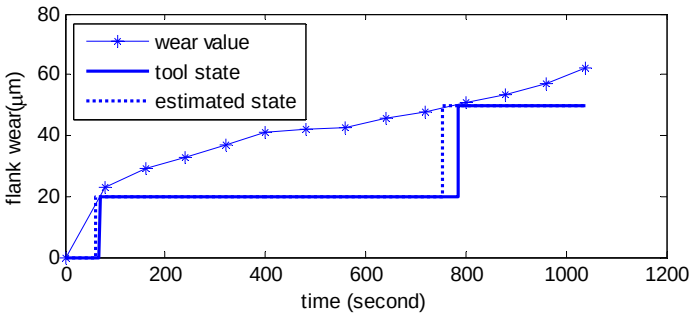


Figure 7.11 b) The left-right HMMs for TCM

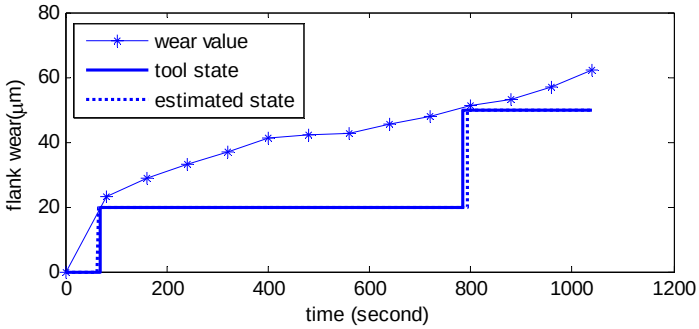
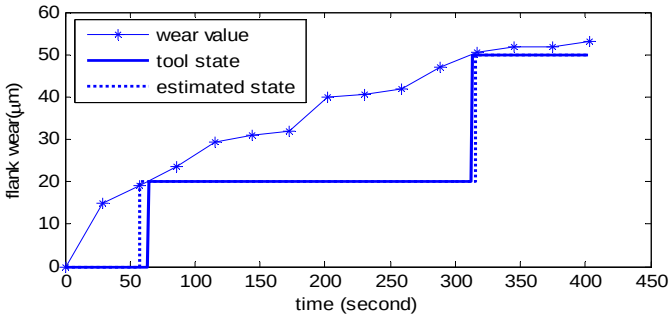


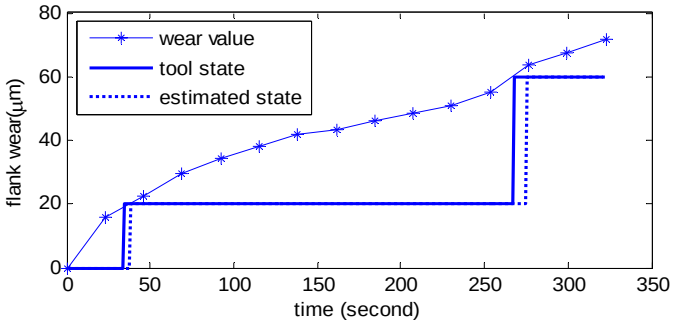
Figure 7.11 c) The medium filtered left-right HMMs for TCM

Figure 7.11 HMMs algorithm improvement

Case studies are shown in figure 7.12 for the data set from tests 16 on copper and also test 27 on steel.



Test 16, Copper
Spindle speed: 18,000 rpm DOC: 150μm
Radial DOC: 225μm
Feed rate: 150mm/min



Test 27, Steel T4
Spindle speed: 18,000 rpm DOC: 60μm
Radial DOC: 80μm
Feed rate: 100mm/min

Figure 7.12 HMMs tool states recognition

7.5 Generalization of the HMM-based Algorithm for TCM

The above discussion assumes that the tool has three states: Slight wear, Medium wear and Severe wear. As a matter of fact, this choice is based on the similarity measurement of signal patterns of different wear values. It may be generalized to include more tool states, which emit features that represent corresponding states while

differentiate it from others. It can be easily implemented with the HMMs - just train these states with separate HMMs and add them to the current HMMs. The augmented HMMs can then estimate more states.

Another potential is the application of HMMs for prognostics. Note that there is a by-product of the Viterbi Coding whereby we estimate the possibilities of the three states while only choosing the highest one. We may apply this for our confidence measurement.

$$\text{Confidence of state } C_i = \frac{p_i}{p_1 + p_2 + p_3} \times 100\%, \ (i = 1, 2, 3) \qquad (7.49)$$

For example, if our confidence of the severe wear state exceeds a certain level (e.g. 80%), we can be more certain predict that the tool will wear out and lose effect soon.

7.6 Conclusion

In this chapter, an approach based on noise-robust continuous HMM for tool wear multi-category classification in micro-milling is proposed. In the HMM, the observations are obtained from the most discriminant features according to their FDR scores to improve HMMs performance for TCM. To overcome the drawbacks of premature state changing due to noisy signal in the left-right HMMs, the state sequence evolution is constrained and decision of the states are made with a medium filter after Vertibi estimation. The effectiveness of the approach has been evaluated under different conditions, e.g. different working conditions, different work piece materials, and variations of observation sequence length. The experimental results indicate that an average recognition rate of as high as 92.5% (ranging from 84.0%-97.3%) and 90.5% ranging from: 84.6%-95.6%) can be achieved for copper and steel respectively. Therefore, the approach based on continuous HMM proposed is highly effective for micro-milling tool wear monitoring. It can be generalized and implemented to include more states for TCM and can also be used for prognostics of tool life.

Chapter 8
Flank Wear Estimation with Continuous HMM

8.1 Experiment setup

Experiments were conducted to verify the application of the HMM algorithm for micro-milling tool wear state estimation. The machine used in the experiments is a MAKINO V55 vertical milling machine driven by a 22kw spindle drive motor. The cutting force was measured with a Kistler 9256A 3-channel dynamometer where the workpiece was mounted (figure 8.1). Tool wear was measured using the Olympus Toolmakers microscope (213 times enlargement). The cutting force output was recorded on a Sony digital tape recorder. The sampling rate of the cutting forces is 6,000 Hz. The force signals were processed within a segment window of 360 points, equivalent to 20 revolutions and 0.06 second for 20,000rpm, and 18 revolutions and 0.06 second for 18,000rpm.

Figure 8.1 Experiment setup

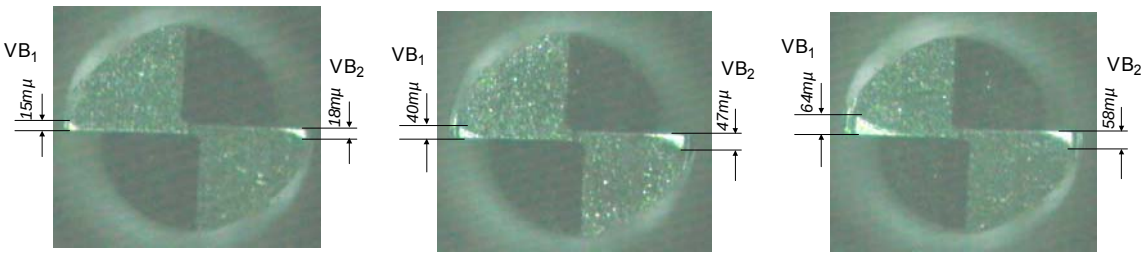
The micro-machining conditions are listed in table 8.1. In total 27 test experiments were conducted with different spindle speed, depth of cut, radial depth of cut, and feed rate for *Φ500μm* and *Φ800μm* tools. The materials used are either copper or steel.

Table 8.1 Experiment setup

Test	Material	Tool diameter (μm)	Spindle Speed (rpm)	Depth of cut (μm)	Radial depth of cut (μm)	Feed rate (mm/min)
1	Copper	500	20,000	30	75	80
2	Copper	500	20,000	30	75	120
3	Copper	500	20,000	40	75	100
4	Copper	500	20,000	60	75	120
5	Copper	500	20,000	60	75	150
6	Copper	500	20,000	80	75	150
7	Copper	500	20,000	80	75	180
8	Copper	500	18,000	60	75	120
9	Copper	500	18,000	80	150	150
10	Copper	500	18,000	100	150	150
11	Copper	500	18,000	100	225	150
12	Copper	500	18,000	100	225	180
13	Copper	500	18,000	120	225	150
14	Copper	500	18,000	100	300	180
15	Copper	500	18,000	120	300	180
16	Copper	500	18,000	150	225	150
17	Copper	500	18,000	120	300	150
18	Steel	800	18,000	80	240	120
19	Steel	800	18,000	80	80	100
20	Steel	800	18,000	80	120	100
21	Steel	800	18,000	80	80	120
22	Steel	800	18,000	80	100	100
23	Steel	800	18,000	80	80	80
24	Steel	800	18,000	60	80	80
25	Steel	800	18,000	60	80	60
26	Steel	800	18,000	60	60	60
27	Steel	800	18,000	60	80	100

8.2 Definition of Tool Wear Level

We took the average of the two flute flank wears $VB=1/2(VB_1+VB_2)$ as the final flank wear value, as demonstrated in figure 8.2**.** In order to establish various wear state for micro-machining, we refer to the Taylor's tool life curve [10]. There are initial wear region, gradual wear region and the accelerated wear regions. In this chapter, the progressive change in the flank wear is approximated by a B-spline curve, and the cross over points of the second order derivative of the Taylor tool life curve is chosen to be the boundary between different regions (figure 8.3). The aforementioned three tool states are defined, and are measured by the threshold that indicates the transition to each state based on the second-order change in the fitted flank wear curve. The average values measured by the thresholds of these three states are listed in table 8.2. We then choose their closest 10μm as the final threshold for convenient representation. This rounding of the threshold is based on the belief that the threshold is roughly estimated while the features can represent their corresponding wear level well.



a) initial wear b) progressive wear c) accelarated wear

Figure 8.2 Flank wear of a micro mill tool with diameter Φ800μm

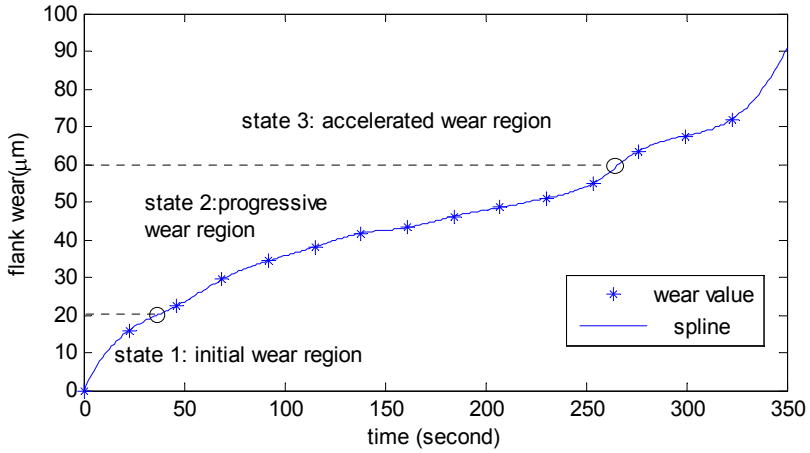


Figure 8.3 tool wear of steel and B-spline interpolation

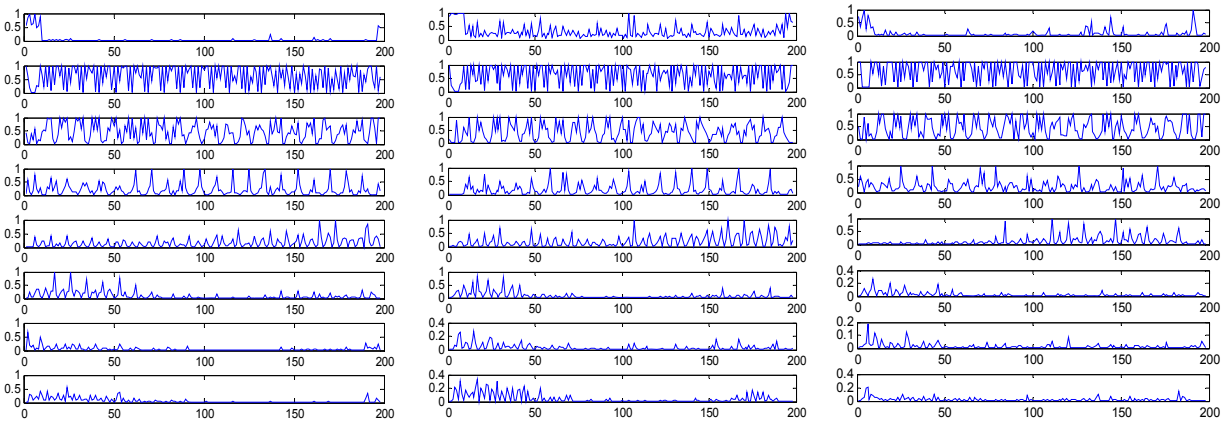
Table 8.2 definition of tool state

Tool diameter	threshold	State 1	State 2	State 3
500 μm	Estimated	17.9μm	17.9-53.8μm	53.8μm
500 μm	Final	0-20μm	20-50μm	>50μm
800 μm	Estimated	23.2μm	23.2-57.5μm	57.5μm
800 μm	Final	0-20μm	20-60μm	>60μm

8.3 Segmentation of Data and Normalization of Feature Vectors

The sampling rate of the cutting forces is 6,000 Hz. The forces are processed with a segment window of 360 points, amount to 20 revolutions and 0.06 second.

For the real force signal, we are confronted with features vectors that vary within different dynamic ranges because of different cutting conditions. For example, cutting forces are much larger when the depth of cut is higher. Thus, features with large values may have a larger influence in the discriminant function than features with small values, although this does not necessarily reflect their discriminability of the tool states. The problem can be overcome by normalizing the features so that their values lie within similar ranges. The extracted 8-component wavelet feature vectors are normalized between [0,1]. Figure 8.4 shows the normalized features from different tool wear level. The different dynamic and distribution properties of features from different tool wear states bring the possibility of tool state estimation. These features are then will trained for building HMMs or input to HMMs for state estimation.

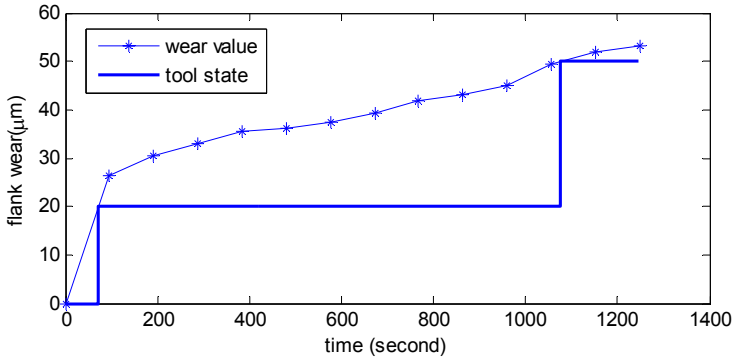


a) initial wear b) progressive wear c) accelerated wear

Figure 8.4 Normalized top 8 discriminant features

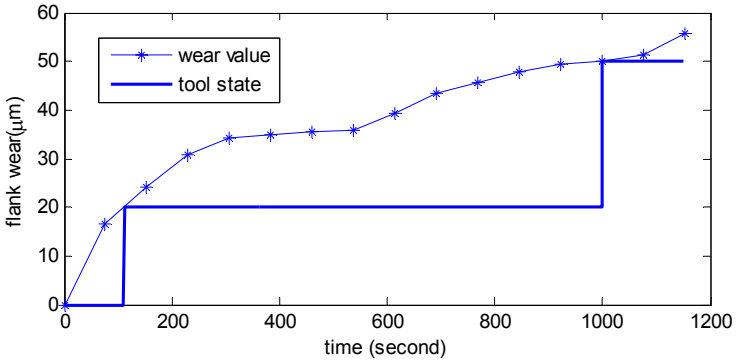
8.4 HMM Training for TCM

The training sets are the test 4, test 10 and test 13 for copper, and test 20 and test 24 for steel. Figure 8.5 -8.7 are the training set for the HMM modeling for micro milling of copper, while Figure 8.8 and figure 8.9 are the selected training set for steel. The selection of the training set is based on the distance to the mean working parameters.



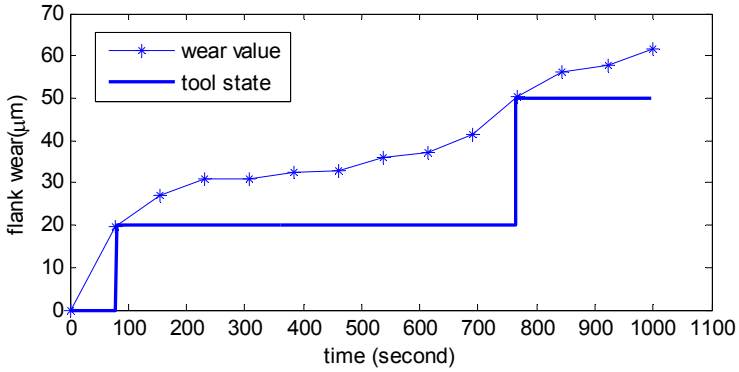
Test 4, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 60μm
Radial DOC: 75μm
Feed rate: 120mm/min

Figure 8.5 Training Test 4



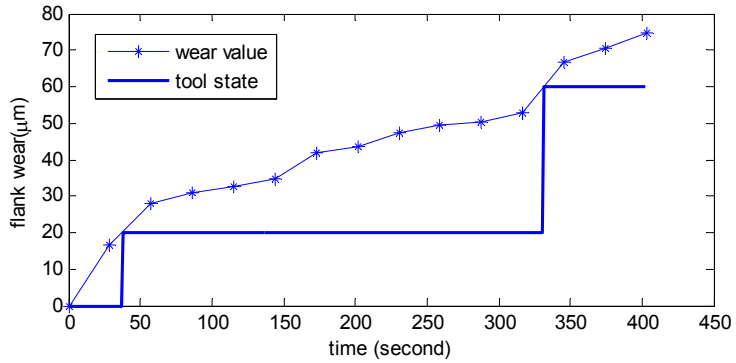
Test 10, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 100μm
Radial DOC: 150μm
Feed rate: 150mm/min

Figure 8.6 Training Test 10



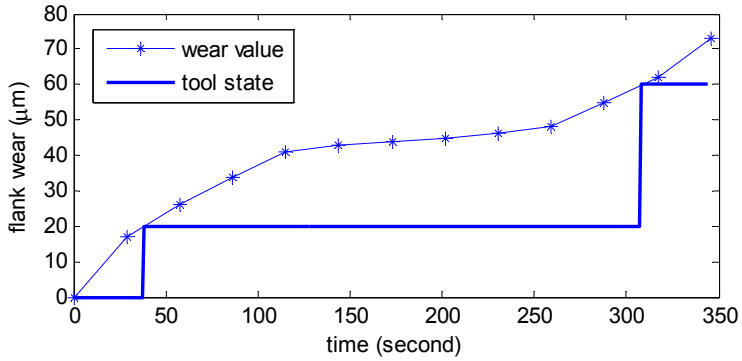
Test 13, Copper
Tool: Φ500 μm
Spindle speed: 18,00 rpm
DOC: 12μm
Radial DOC: 225μm
Feed rate: 150mm/min

Figure 8.7 Training Test 13



Test 20, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 120μm
Feed rate: 100mm/min

Figure 8.8 Training Test 20



Test 24, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 60μm
Radial DOC: 80μm
Feed rate: 80mm/min

Figure 8.9 Training Test 24

8.5 HMM for Tool State Estimation

Because there are 5 HMMs trained, we need to decide which HMMs suit best for the tool state estimation. Given the working condition, a similarity measure based on the Euclidian distance of the working condition is first determined as follows:

$$d_i = \left[\left(\frac{v - v_{ri}}{v_{ri}}\right)^2 + \left(\frac{f - f_{ri}}{f_{ri}}\right)^2 + \left(\frac{d_v - d_{vri}}{d_{vri}}\right)^2 + \left(\frac{d_r - d_{rri}}{d_{rri}}\right)^2 \right]^{1/2} \qquad (7.45)$$

where d_i is the distance, v_{ri} is the *i-th* reference spindle speed, f_{ri} is the *i-th* reference feed rate, d_{ri} is the *i-th* reference vertical depth of cut, and d_{rri} is the *i-th* reference radial depth of cut. Our aim is to find the one with the minimum distance, $\min d_i$. The other two variables, i.e. work piece materials and tool diameters, are not included here as they are usually predetermined and assumed to be the same.

The unknown state is then uncovered with the model given the maximum probability of the observed features:

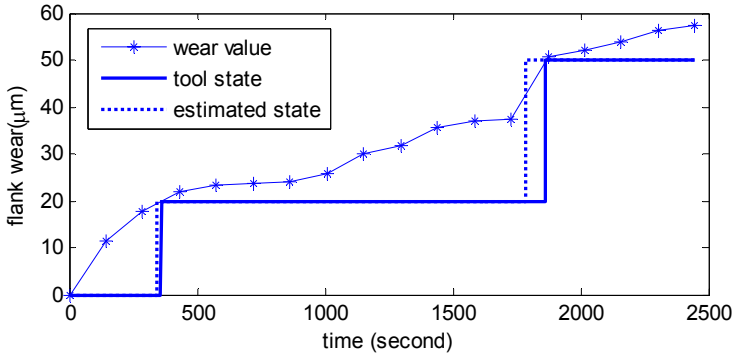
$$w^* = \arg\max p_w, w = 1, 2, ... \quad (7.46)$$

Thus, w* is the optimum class for the observation ***Y***.

Figure 10 - figure 23 are the HMM tool state recognition result s for the micro milling of copper with tool diameter Φ500 μm. The corresponding classification rates are listed in table 8.3 .

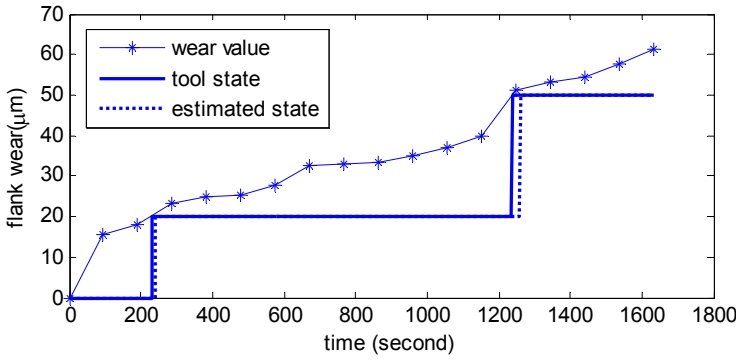
Table 8.3 Classification rate of copper with 3- state HMMs

Classification rate of the different tests (%)														
Test 1	Test 2	Test 3	Test 5	Test 6	Test 7	Test 8	Test 9	Test 11	Test 12	Test 14	Test 15	Test 16	Test 17	Aver a-ge
89.5	95.3	96.7	88.5	95.3	89.4	96.8	94.1	97.3	96.5	84.0	93.8	90.1	87.2	92.5



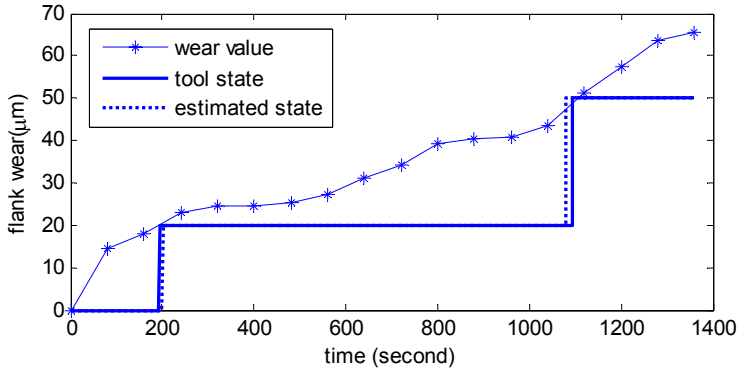
Test 1, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 30μm
Radial DOC: 75μm
Feed rate: 80mm/min

Figure 8.10 HMMs tool states recognition Test 1



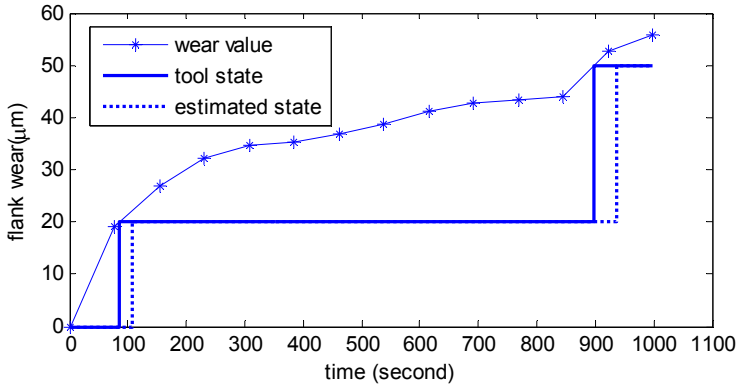
Test 2, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 30μm
Radial DOC: 75μm
Feed rate: 120mm/min

Figure 8.11 HMMs tool states recognition Test 2



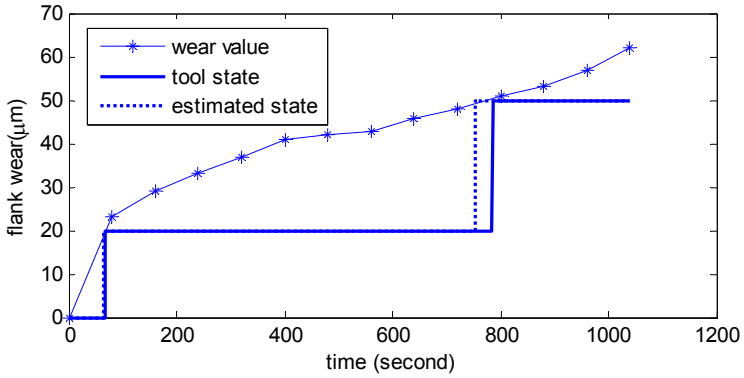
Test 3, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 40μm
Radial DOC: 75μm
Feed rate: 100mm/min

Figure 8.12 HMMs tool states recognition Test 3



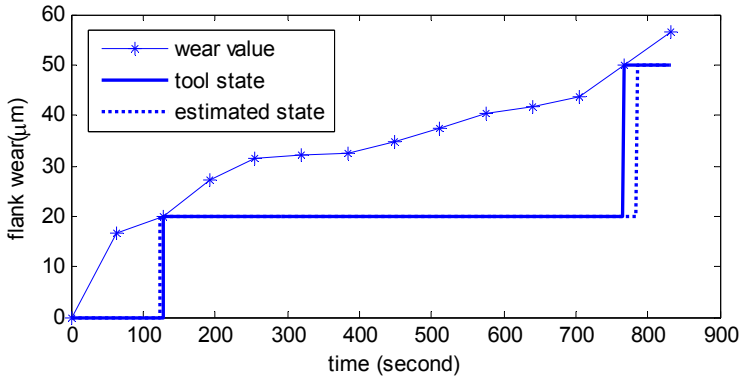
Test 5, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 60μm
Radial DOC: 75μm
Feed rate: 150mm/min

Figure 8.13 HMMs tool states recognition Test 5



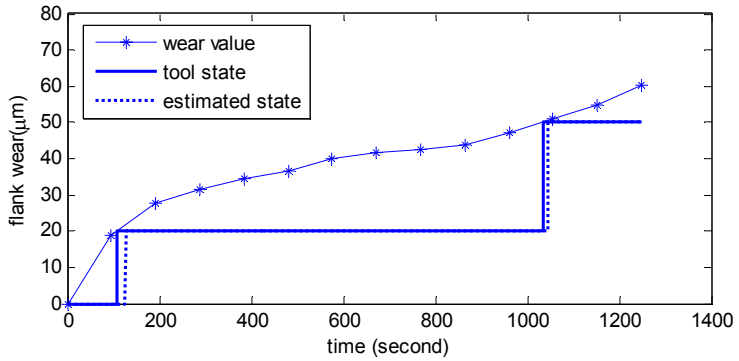
Test 6, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 80μm
Radial DOC: 75μm
Feed rate: 150mm/min

Figure 8.14 HMMs tool states recognition Test 6



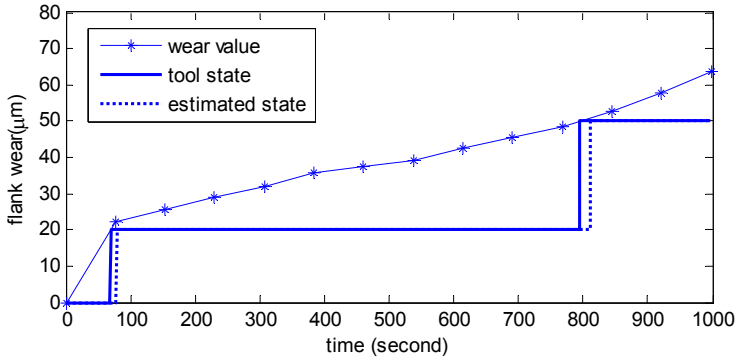
Test 7, Copper
Tool: Φ500 μm
Spindle speed: 20,000rpm
DOC: 80μm
Radial DOC: 75μm
Feed rate: 180mm/min

Figure 8.15 HMMs tool states recognition Test 7



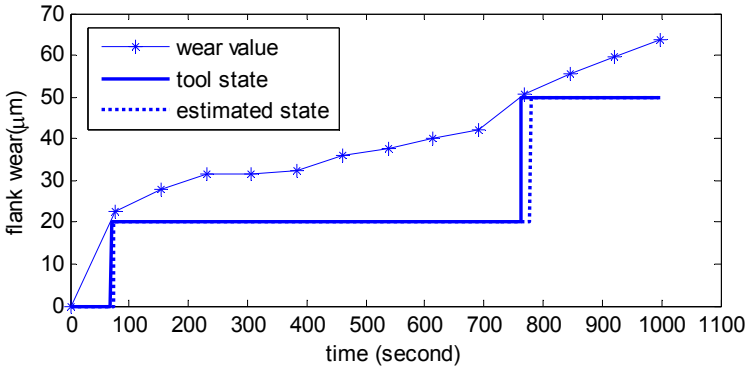
Test 8, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 60μm
Radial DOC: 75μm
Feed rate: 120mm/min

Figure 8.16 HMMs tool states recognition Test 8



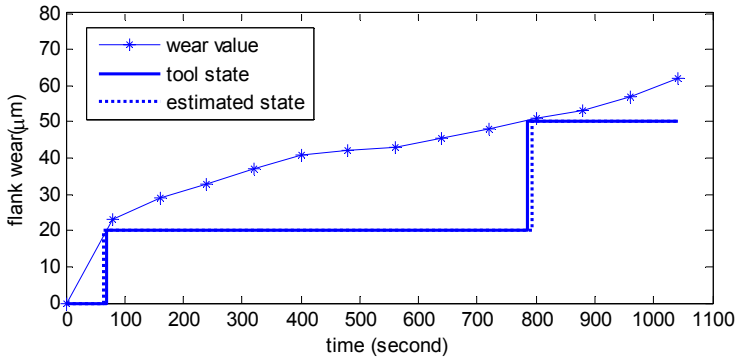
Test 9, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 150μm
Feed rate: 150mm/min

Figure 8.17 HMMs tool states recognition Test 9



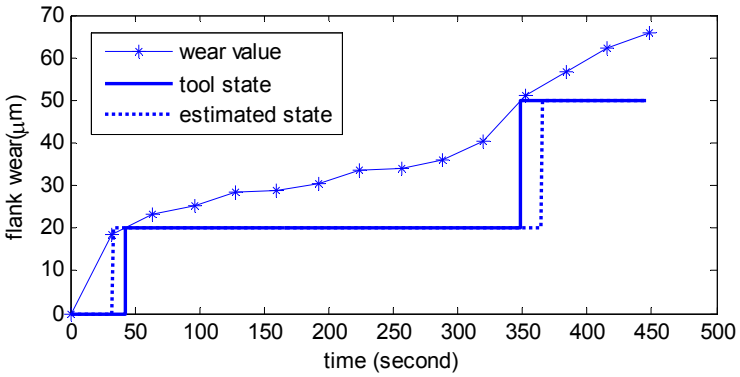
Test 11, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 100μm
Radial DOC: 225μm
Feed rate: 150mm/min

Figure 8.18 HMMs tool states recognition Test 11



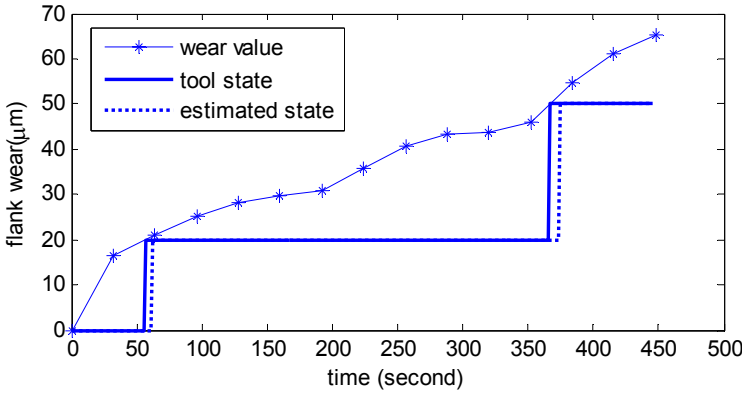
Test 12, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 100μm
Radial DOC: 225μm
Feed rate: 180mm/min

Figure 8.19 HMMs tool states recognition Test 12



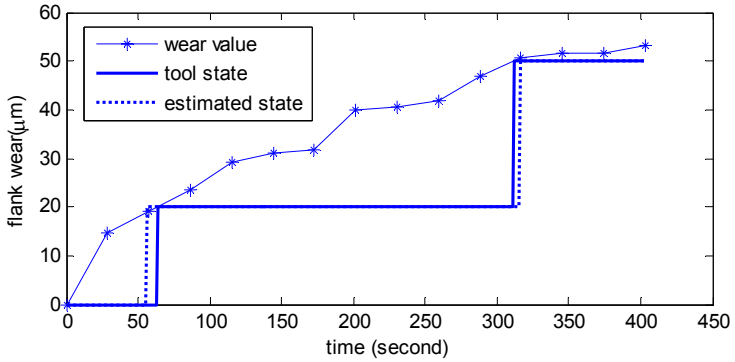
Test 14, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 100μm
Radial DOC: 300μm
Feed rate: 180mm/min

Figure 8.20 HMMs tool states recognition Test 14



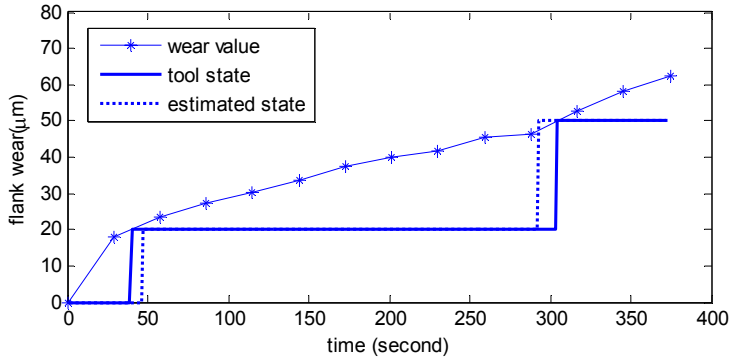
Test 15, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 120μm
Radial DOC: 300μm
Feed rate: 180mm/min

Figure 8.21 HMMs tool states recognition Test 15



Test 16, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 150μm
Radial DOC: 225μm
Feed rate: 150mm/min

Figure 8.22 HMMs tool states recognition Test 16



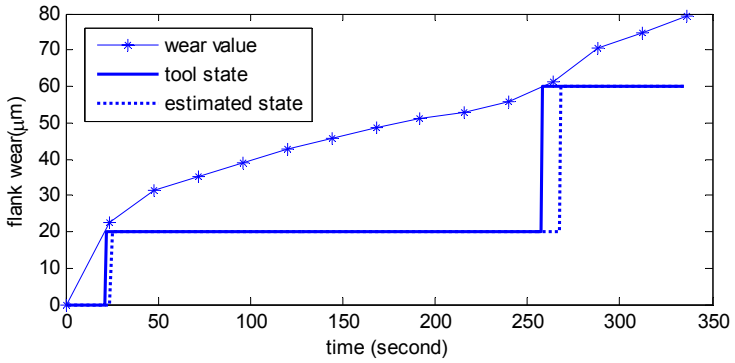
Test 17, Copper
Tool: Φ500 μm
Spindle speed: 18,000rpm
DOC: 150μm
Radial DOC: 300μm
Feed rate: 150mm/min

Figure 8.23 HMMs tool states recognition Test 17

Figure 24 - figure 31 are the HMM tool state recognition result s for the micro milling of steel with tool diameter Φ800 μm. The corresponding classification rates are listed in table 8.4.

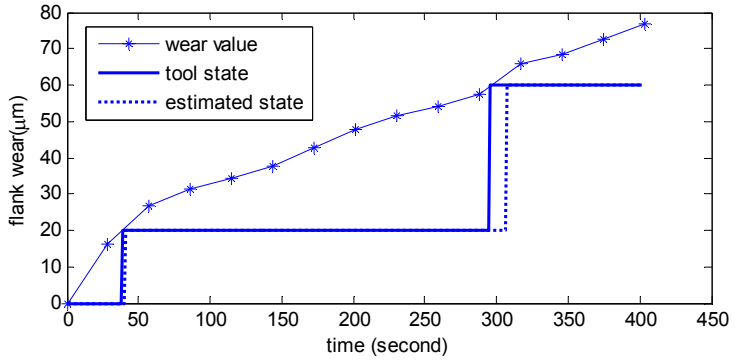
Table 8.4 Classification rate of steel with 3-state HMMs

Classification rate of the different tests (%)								
Test 18	Test 19	Test 21	Test 22	Test 23	Test 25	Test 26	Test 27	Average
88.3	90.2	89.9	84.6	91.2	95.6	93.7	89.5	90.4



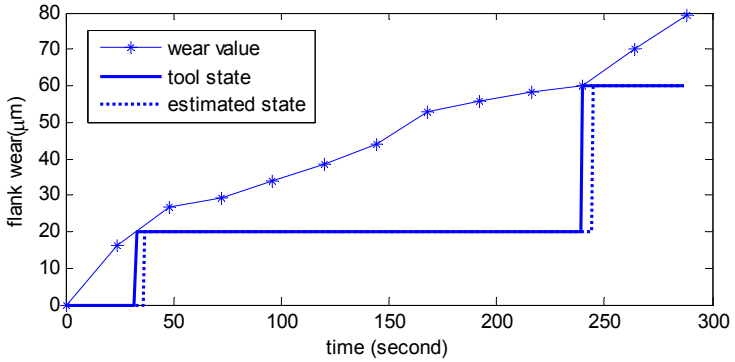
Test 18, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 240μm
Feed rate: 120mm/min

Figure 8.24 HMMs tool states recognition Test 18



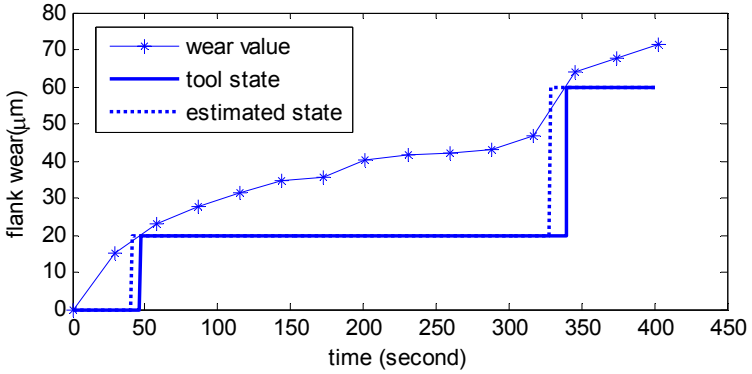
Test 19, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 80 μm
Feed rate: 100mm/min

Figure 8.25 HMMs tool states recognition Test 19



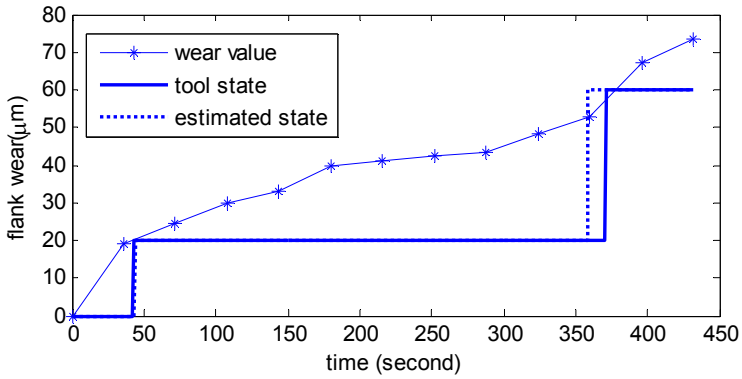
Test 21, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 80 μm
Feed rate: 120mm/min

Figure 8.26 HMMs tool states recognition Test 21



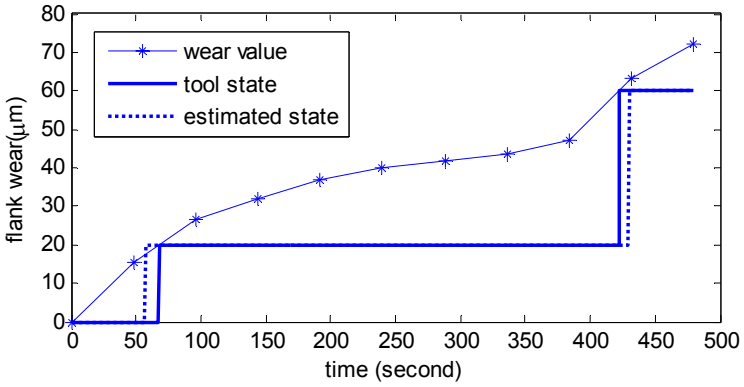
Test 22, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 1080 μm
Feed rate: 100mm/min

Figure 8.27 HMMs tool states recognition Test 22



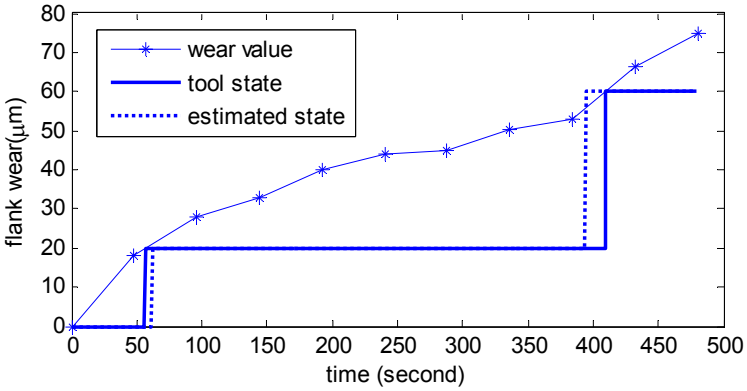
Test 23, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 80μm
Radial DOC: 80μm
Feed rate: 80mm/min

Figure 8.28 HMMs tool states recognition Test 23



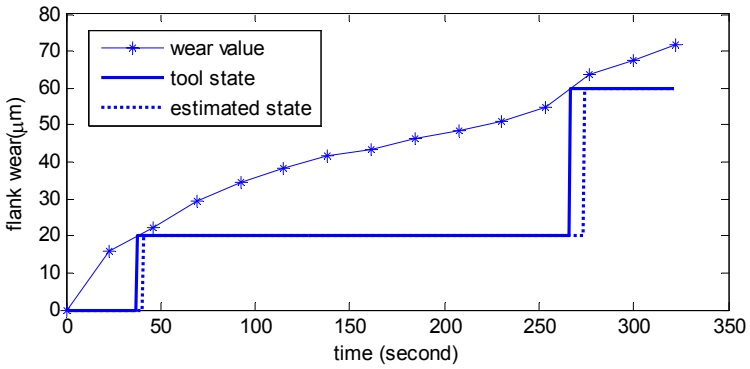
Test 25, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 60μm
Radial DOC: 80μm
Feed rate: 60mm/min

Figure 8.29 HMMs tool states recognition Test 25



Test 26, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 60μm
Radial DOC: 60μm
Feed rate: 60mm/min

Figure 8.30 HMMs tool states recognition Test 26



Test 27, Steel T4
Tool: Φ800 μm
Spindle speed: 18,000rpm
DOC: 60μm
Radial DOC: 80μm
Feed rate: 100mm/min

Figure 8.31 HMMs tool states recognition Test 27

8.6 Results and Conclusion

In this chapter, the effectiveness of the approach was evaluated under different conditions, e.g. different working conditions, different work piece materials, and variations of observation sequence length. In the HMM, the observations are obtained from the most discriminant features according to their FDR scores to improve HMMs performance for TCM. To overcome the drawbacks of premature state changing due to noisy signal in the left-right HMMs, the state sequence evolution is constrained and decision of the states is made with a medium filter after Vertibi estimation. With the top 8 discriminant features from both time and wavelet domain, the experimental results showed that the HMMs achieved an average recognition rate of as high as 92.5% (ranges: 84.0%-97.3%) and 90.5% (ranges: 84.6%-95.6%) for copper and steel respectively. Therefore, the approach based on continuous HMMs is highly effective for micro-milling tool wear monitoring. The approach can be generalized and adapted to include more states for TCM and can also be used for prognostics of tool life as discussed in section 7.5.

Chapter 9

Conclusion and Recommendations for Future Work

Tool condition monitoring in micro machining poses new challenges compared to conventional machining, due to the high tool wear rate and high precision requirement associated with the dimensions to be produced at micro-level. In this thesis, a multi-category classification approach is proposed for tool wear monitoring based on HMMs. In the framework of HMMs, the cutting forces are denoised first, and then the features of the cutting forces are selected from both time and wavelet domain based on their discriminant scores. These features are iteratively learned by the HMMs to map the relationships between force features and tool wear states. Different tool wear states are modeled as separate HMMs. The tool wear state is then classified by the HMM that has the maximum probability to indicate the test features.

9.1 Contributions

The contributions in the thesis are the solution methods to the three main problems of micro-milling TCM.

1. Cutting force Denoising in Case of Non-Gaussian Noise

The signal noise is correlated to the machining conditions and has wide harmonics and non-Gaussian distribution. The traditional denoising methods, based on Gaussian noise assumption, are not effective because the noise is found to contain high non-Gaussian component in the experiment. The proposed approach employs fixed-point independent component analysis (FastICA). It assumes that both the Gaussian noise and non-Gaussian noises are signal sources and to be instantaneous mixtures of sources The signal denoising problem is treated as blind source separation (BSS), under the measure of source's negnentropy. The results show that FastICA separates both Gaussian and non-Gaussian noise sources. It is superior to signal denoising with wavelet thresholding, which only separate Gaussian noise. Wavelet thresholding estimates too small threshold in this case, and does not work well by discarding smaller coefficients since the non-Gaussian noise coefficients are not small and evenly

distributed in certain bands. The reason behind this difference is also verified. This approach broadens the ideas of the BSS in communication by solving the BSS as a signal denoising process.

2 Discriminant Feature Dimension Reduction

Fisher's linear discriminant analysis (FDA) is adapted for this purpose. In the discriminant selection, features are chosen to maximize class separation and are ranked by their separation ability between different classes. Other popular feature dimension reduction methods, such as principal component analysis (PCA) and automatic relevance determination (ARD), are also discussed and compared with the discriminant method with their classification rate. The reasons that the FDA is superior to both PCA and ARD in feature selection are also discussed. The effectiveness of the approach was evaluated under different working conditions. Experimental results indicate that the selection of the top 8 most discriminant features can achieve the highest classification, with average classification rate (for 3-state tools) being as high as 97.2% can be achieved. Therefore, the FDR approach is a successful feature selection tool for HMMs modeling in TCM. On the other hand, unlike the other methods such as mutual information, the FDR method is robust because there are no assumptions made with regard to the features' distributions. Features from both time and wavelet domain are ranked only on their class-separation ability. Selecting the most discriminant features according to their FDR scores improve the performance of the HMMs for TCM.

3. Noise-robust Multi-category Tool Wear State Estimation

The tool states are defined into three states with Tayor's reference model, using second order difference variations to indicate tool state changes. Continuous Hidden Markov models (HMMs) are adapted for stochastic modeling of the tool wear in micro-milling, and estimation of the tool wear state based on the cutting force features. The continuous HMMs overcome the problem of feature degrading, such as that in discrete HMMs. But in the HMM, the state transition matrix is constrained to only change right but not in reverse, as some time may be misclassified the state because of signal noise or signal processing noise and leading the current signal patterns to be more similar to the later tool state. To overcome the drawbacks of premature state changing due to noisy signal in the left-right HMMs, the state sequence evolution is constrained and

decision of the states are made with a medium filter after Vertibi estimation. This medium state is then the final tool wear state. The effectiveness of the approach has been evaluated under different conditions, e.g. different working conditions, different work piece materials, and variations of observation sequence length. The experimental results indicate that an average recognition rate of as high as 92.5% (ranging from 84.0%-97.3%) and 90.5% ranging from: 84.6%-95.6%) can be achieved for micro-milling of copper and steel respectively (Refer to appendix A for linear regression and neural networks for tool state estimation). Therefore, the proposed continuous HMM is highly effective for micro-milling tool wear monitoring. It can be generalized and implemented to include more states for TCM and can also be used for prognostics of tool life.

9.2 Recommendations for Future Work

In the area of TCM in micro-milling there are five major areas where the research mentioned in this thesis can be expanded.

1. Cutting Dynamic Investigation

The cutting mechanism in the micro machining of materials, while not presented in this thesis, are an important and promising area of research. The theory of the cutting operation has yet to enable approached prediction of high speed cutting process.

2. Micro-milling TCM in Free-form Surface Generation

One of the most attractive advantages of micro-milling is the free-form 3D micro structures machining, but we only concentrate on constant depth of cut end milling in this study. It is challenging for TCM since the working conditions changes continuously.

3. Application of HMMs for Prognostics

Note that there is a by-product of the Viterbi Coding for the estimation the possibilities of the three states while only choosing the highest one. This may be applied for confidence measurement. For example, if the confidence of the severe wear state exceeds a certain level (e.g. 80%), we can be more certain in predicting when the

tool will wear out and lose its effect soon. The difficulty is on the selection of the suitable threshold.

4. Involve More States in the HMMs

It is convenient to include more tool states in HMMs for TCM. By just adding more features from the corresponding states for training, the other trained states and features will not be changed.

However, it should be cautious to choose the state level definition, as more states do not always mean better classification, and sometime more states are not necessary. Figure 9.1 illustrate the idea of involving 7 states, with 10 μm increment for each state. The HMMs always pre-enter the state in the estimation, and the classification rate is low. It was found that even grouping into groups of (states 1, 2), states (3, 4, 5), and (state 6, 7) as the former three categories, the classification rate was still lower than the former approach.

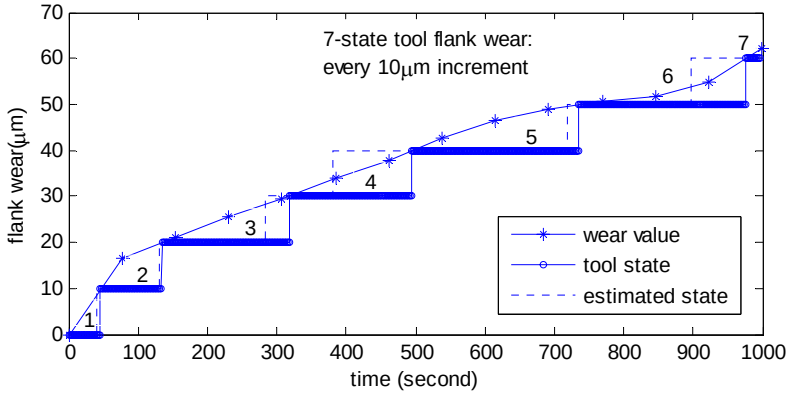


Figure 9.1 The seven categories HMMs for TCM

Reference

[1]. Abramovich, F. & Sapatinas, T. (1999). Bayesian approach to wavelet decomposition and shrinkage. In Bayesian Inference in Wavelet Based Models, M¨uller, P. & Vidakovic, B. (Eds.), Lecture Notes in Statistics, 141, pp. 33–50, New York: Springer-Verlag.

[2]. Abramovich, F., Amato, U. & Angelini, C. (2004), On optimality of Bayesian wavelet estimators, Scandinavian Journal of Statistics, 31, 217–234.

[3]. Abramovich, F., Bailey, T.C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. The Statistician, 49, 1–29.

[4]. Abu-Mahfouz I., Drilling wear detection and classification using vibration signals and artificial neural network, International Journal of Machine Tools & Manufacture, v 43, n 7, May 2003, p 707-20

[5]. Ali A. H., and Elijah K.A.J., statistical process control for cutting tool of acoustic emission monitoring, Mechanical Systems and Signal Processing (1989) 3(4), 405-424

[6]. Ali A. H., Elijah K.A. J. and Gray D. H., A dynamic model for tool detection using acoustic emission, Mechanical Systems and Signal Processing, Volume 9, Number 4, 1995, pp. 415-428.

[7]. Altintas Y., Prediction of cutting forces and tool breakage in milling from feed drive current measurement, ASME J. Eng. Ind., Vol. 114, 386-392, 1992

[8]. Altintas, Y., and Yellowley, I., The process detection of tool failure in milling using cutting force models, ASME Journal of Engineering for Industry 111 (1989) 149–157.

[9]. Altintas, Y., In-process detection of tool breakages using time series monitoring of cutting forces, Int. J. Mach. Tools Manufacture. 28 (2) (1988) 157-172.

[10]. Altintas, Y., Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design, Cambridge University Press, 2000.

[11]. Altmann, J., and Mathew, J., Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnostics, Mechanical Systems and Signal Processing, 15 (2001) 963–977.

[12]. Antoniadis, A., Leporini, D. and Pesquet, J.C., Wavelet thresholding for some classes of non-Gaussian noise, Statistica Neerlandica (2002) Vol. 56, (4), pp. 434–453

[13]. Antoniadis, A. and Gijbels, I. (2000) Detecting abrupt changes by wavelet methods. Journal of Nonparametric Statistics, 2002, Vol. 14(1-2), pp. 7–29.

[14]. Antonio G. V. J., Juan A.N.F, Ruben M.M., L. Enrique Sucar, and Ciro A. R, Tool-Wear Monitoring Based on Continuous Hidden Markov Models, M. Lazo and A. Sanfeliu (Eds.): CIARP 2005, LNCS 3773, pp. 880–890, 2005, Springer-Verlag Berlin Heidelberg 2005

[15]. Atlas, L., Ostendorf, M., and Bernard, G., Hidden Markov models for monitoring machining tool-wear, Proc. of ICASSP, 3887--3890, 2000.

[16]. Averkamp, R. and Houdr, C., Wavelet Thresholding for Non-Necessarily Gaussian Noise: Idealism, Annals of Statistics 31, 110-151 (2003).

[17]. Bailey, T. C., Sapatinas, T., Powell, K. J. and Krzanowski, W. J. (1998) Signal detection in underwater sound using wavelets. J. Am. Statist. Ass., 93, 73-83.

[18]. Bao, W.Y., and Tansel, I.N. Modeling Micro-end-milling operations. Part III: Influence of Tool Wear, International Journal of Machine Tools & Manufacture, Vol. 40, pp. 2193-2211.

[19]. Bao, W.Y., and Tansel,I.N., Modeling micro-end-milling operations. Part I: analytical cutting force model, Int. J. Mach. Tools Manufacture Vol. 40 pp.2155–2173, 2000

[20]. Baruah, P., and Chinnam, R. B., HMMs for diagnostics and prognostics in machining processes, International Journal of Production Research, Vol. 43, No. 6, 15 March 2005, 1275–1293.

[21]. Bell, A.J.and Sejnowski, T.J., An information-maximisation approach to blind separation and blind deconvolution. Neural computation, 7(6):1004–1034, 1995.

[22]. Berger,B. S., Minis, I., Harley, J., Rokni, M. and Papadopoulos, M., wavelet based cutting state identification, Journal of Sound and Vibration (1998) 213(5) 813-827

[23]. Bilmes, J., What HMMs Can Do? Technical Report, University of Washington, 2002.

[24]. Boothroyd, G., and Knight, W. A., Fundamentals of Machining and Machine Tools, Redwood City, CA: Addison-Wesley, 1989.

[25]. Bukkapatnam, S.T.S., Kumar, S.R.T., Lakhtakia, A., and Srinivasan, P., The neighborhood method and its coupling with the wavelet method for signal separation of chaotic signals, Signal Processing 82 (2002) 1351–1374.

[26]. Byrne, G., Dornfeld, D., Inasaki, I., Konig, W., and Teti, R., Tool condition monitoring (TCM)—the status of research and industrial application, Annals of CIRP 44 (2) (1995) 541–567.

[27]. Byrne, G., Dornfeld D., Denkena, B., Advancing Cutting Technology, Annals of CIRP Keynotes, 2003.

[28]. Cardoso, J. F., Blind signal separation: statistical principles. 1998 Proceedings of the IEEE, Vol 9, No. 10, pp 2009-2025, Oct. 1998.

[29]. Cetin, O. and Ostendorf, M., Multi-rate Hidden Markov Models and their Application to Machine Tool-Wear Classification. In The 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), pp 837-840, 2004.

[30]. Chae J, Park S S and Freiheit T, Investigation of micro-cutting operations. International Journal of Machine Tools & Manufacture, 2006 (46): 313-332

[31]. Chen, J.C., and Chen W.L., Tool breakage detection system using an accelerometer sensor. Journal of Intelligent Manufacturing. 10(2): p. p 187-197, 1999

[32]. Chipman H.A., Kolaczyk E.D. and McCulloch R.E. Adaptive Bayesian Wavelet Shrinkage. J. Am. Statist. Ass., 92 (1997)., 1413–1421.

[33]. Choi Y., Narayanaswami R., and Chandra A., Tool wear monitoring in ramp cuts in end milling using the wavelet transform, International Journal of Advanced Manufacturing Technology, v 23, 2004, p 419-28.

[34]. Chou, K. C. and Heck, L.P. A Multiscale Stochastic Modeling Approach To The Monitoring Of Mechanical Systems, IEEE 1994.

[35]. Chu, W., Keerthi, S.S., Ong, C. J. and Ghahramani, Z., (2005) Bayesian support vector machines for feature ranking and selection, In I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, editors, Feature Extraction, Foundations and Applications, Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, 2006.

[36]. Comon, P. Independent component analysis: a new concept? Signal Processing, 36: 287-314, 1994.

[37]. Daubeches, I., Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[38]. Daubechies, I.: The wavelet transformation, time-frequency localization and signal analysis. IEEE Trans. Inform. Theory 36 (1990), p. 961- 1005.

[39]. Diei, E.N., and Dornfeld, D.A., A model of tool fracture generated acoustic emission during machining, ASME J. of Eng. for Industry, 109 (1987) 227-233

[40]. Derrico, G. E., 1997, A Systems Theory Approach to Modeling of Cutting Temperature with Experimental Identification, International Journal of Machine Tools & Manufacture 37 (1997), pp. 149–158.

[41]. Dimla S.D.E., and Lister, P.M., On-line metal cutting tool condition monitoring I: force and vibration analyses, International Journal of Machine Tools & Manufacture 40 (2000) 739–768.

[42]. Dimla S.D.E., and Lister, P.M., On-line metal cutting tool condition monitoring II: tool-state classification using multi-layer perceptron neural networks, International Journal of Machine Tools & Manufacture 40 (2000) 769–781.

[43]. Donoho, D. L., and Johnstone, I. M. Adapting to Unknown Smoothness via Wavelet Shrinkage, Journal of the American Statistical Association, (1995), 90(432), 1200-1224.

[44]. Donoho, D. L., De-noising by soft-thresholding. IEEE Transactions on Information Theory, 41(3):613–627, May 1995.

[45]. Donoho, D.L. and Johnstone, I.M. Minimax estimation via wavelet shrinkage. Ann. Statist., 26(1998), 879–921.

[46]. Dowling, M.J., Application of non-stationary analysis to machinery monitoring, ICASSP-93. 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993, p 59-62 vol.1.

[47]. Du, R., Elbestawi, M. A. and Wu, S.M., , Computer automated monitoring of manufacturing processes: Part 2, applications, Transactions of ASME, Journal of Engineering for Industry, May, 1995.

[48]. Du, R., Elbestawi, M. A. and Wu, S.M., Computer automated monitoring of manufacturing processes: Part 1, monitoring decision-making methods, Transactions of ASME, Journal of Engineering for Industry, May, 1995.

[49]. Duda, R.O., Hart. P. E. and Stork, D. G., Pattern Classification, (2nd ed.), Wiley Interscience, 2001.

[50]. Ehmann, K.F., Devor, R. E., and Kapoor, S.G., Micro/Meso-scale Mechanical Manufacturing Manufacturing Opportunities and Challenges-, Proceedings, JSME/ASME International Conference on Materials and Processing, Ocober 15-18, Honolulu, HI, Vol. 1 pp. 6-13, 2002

[51]. Elbestawi, M.A., Papazafiriou, T.A. and Du, R.X. In-process Monitoring of Tool Wear in Milling Using Cutting Force Signature, International Journal of Machine Tools & Manufacture, Vol. 31, No. 1, pp. 55-73. 1991.

[52]. El-Wardany, T. I., Gaoand, D. and Elbestawi, M. A., Tool condition monitoring in drilling using vibration signature analysis, International Journal of Machine Tools and Manufacture, Volume 36, Issue 6 , June 1996, Pages 687-711.

[53]. Ernst M K, Dmitri A R, Tatyana N B and Semion A T, Micromechanical engineering: a basis for the low-cost manufacturing of mechanical microdevices using micro equipment, J. Micromechanics. Microengineering. 6 (1996) 410–425.

[54]. Errtunc, H. M., Looparo, K. A. and Ocak, H., Tool Wear Condition Monitoring in Drilling Operations Using Hidden Marko Modes, International Journal of Machine Tools & Manufacture, Vol. 41, pp. 1363-1384. 2001.

[55]. Ertunc, H. M. and Oysu, C. Drill wear monitoring using cutting force signals, Mechatronics 14 (2004) 533–548.

[56]. Fish, R. K., Ostendorf, M., Bernard, G. D., and Castanon, D. A., Multilevel Classification of Milling Tool Wear with Confidence Estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 1, January 2003.

[57]. Fraley, C. and A. E. Raftery (1998). How many clusters? Which clustering method? -Answers via model-based cluster analysis. The Computer Journal 41, 578-588.

[58]. Friedrich, C.R. and Coane, P.J., Micromilling Development and Applications for Microfabrication, Journal of Microelectronic Engineering (35), 1997, pp.367-372.

[59]. Fu, J. C., Troy, C. A. and Mori, K., The prediction of small drill bit breakage by wavelet based joint time-scale analysis. CIRP Proceedings- Manufacturing Systems, 1996, 25, 385-389.

[60]. Gabor, D., Theory of communication, J. Inst. Elec. Eng. 93 (III), 429-457 (1946).

[61]. Gao H.Y. and Bruce A.G. WaveShrink with firm shrinkage. Statistica Sinica, 7(1997), 855–874

[62]. Gelle, G., and Colas, M., Blind source searation: A tool for rotating machine monitoring by vibration analysis? Journal of Sound and Vibration Vol. 248(5), pp.865-885, 2001

[63]. Ghasempoor, A., Jeswiet, J. and Moore, T.N., Real time implementation of on-line tool condition monitoring in turning, International Journal of Machine Tools and Manufacture, Volume 39, Issue 12 , December 1999, Pages 1883-1902

[64]. Ghasempoor, A., Jeswiet, J., and Moore, T.N., Real time implementation of on-line tool condition monitoring in turning, International Journal of Machine Tools & Manufacture 39 (1999) 1883–1902

[65]. Goebel, K. and Yan, W., Feature Selection for Tool Wear Diagnosis Using Soft Computing Techniques, GE Technical Information Series, 2001CRD011, February 2001.

[66]. Gong, W., Obikawa, T. and Shirakashi, T. (1997) Monitoring of tool wear states in turning based on wavelet analysis. JSME Int. J. (Series C) 40, 447–453.

[67]. Hambaba, A., and Hu, A. E., Multiresolution error detection on early fatigue cracks in gears, 2000 IEEE Aerospace Conference Proceedings, Vol 6, 367–372.

[68]. Han, D., and Zhang, S.N., Comparison between Windowed FFT and Hilbert-Huang Transform for Analyzing Time Series with Poissonian Fluctuations: A Case Study, Chinese Journal of Astronomy and Astrophysics, In Press

[69]. Hardle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). Wavelets, Approximation and Statistical Applications. Lecture Notes in Statistics, vol. 129. Springer, N.Y.

[70]. Heck, L. P., and McClellan, J. H., Mechanical System Monitoring Using HMMs, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 1697–1700, 1991.

[71]. Heck. L.P. and Chou, K. C., Gaussian Mixture Model Classifiers for Machine Monitoring, ICASSP-94. 1994 IEEE International Conference on Acoustics, Speech and Signal Processing, 1994, pt. 6, p VI/133-6 vol.6.

[72]. Hong, G.S., Rahman, M., and Zhou, Q., Using neural network for tool condition monitoring based on wavelet decomposition, Int. J. Mach. Tools Manufacture, Vol. 36, No. 5, pp. 551-566, 1996

[73]. Huang, N.E., Shen, Z., Long, S.R., MWu,.L.C., Shih, H.H. Zheng, Q.N., Yen, N.C., Tung, C.C., and Liu, H.H., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,

Proceeding of the Royal Society of London Series A—Mathematical Physical and Engineering Sciences 454 (1998) 903–995.

[74]. Huang, P.T., Chen, J.C. and Chou, C.Y., A statistical approach in detecting tool breakage in end milling operations, J. of Industrial Technology, Vol. 15, No. 3, 1999.

[75]. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, 1999, 10, 626-634

[76]. Hyvarinen, A., Karhunen, J., and Oja, E., Independent Component Analysis, 2001 John Wiley & Sons.

[77]. Iannarilli, F.J., and Rubin, P.A., Feature Selection for Multiclass Discrimination via Mixed-Integer Linear Programming, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 779-783, June 2003

[78]. Ikawa, N., Shimada, S., and Tanaka, H., 1992, Minimum Thickness of Cut in Micro-machining, Nanotechnology, 3, pp. 6–9.

[79]. International Standard Organization (ISO 8688-2: 1989) Tool life testing in milling -Part 2: End milling, 1989.

[80]. Fu, J. C., Troy,C. A. and Phillips,P. J., A Matching Pursuit Approach To Small Drill Bit Breakage Prediction, Int. J. Prod. Res., 1999, Vol. 37, No. 14, 3247- 3261.

[81]. Jansen, M., .Noise reduction by wavelet thresholding, Lecture notes in statictics 161, Springer, 2001

[82]. Jansen, M., Wavelet denoising, PhD thesis, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, October 2001.

[83]. Jeong, M. K., Lu, J. C., Huo, X., Vidakovic, B., and Chen, D. Wavelet-Based Data Reduction Techniques for Fault Detection and Classification, Technical Report. The School of Industrial and Systems Engineering, Georgia Institute of Technology, USA. (2002).

[84]. Jolliffe, I. T., Principal Component Analysis. 2nd edition, New York: Springer-Verlag, 2002.

[85]. Jun, M.B., Modeling and Analysis of Micro-End Milling Dynamics, PhD Dissertation, University of Illinois Urbana-Champaign, 2005.

[86]. K. Fukunaga, Introduction to statistical pattern recognition, 2nd ed., Academic Press, 1990.

[87]. Kamarthi, S.V. and Pittner, S., Fast Fourier and Wavelet Transform for Flank Wear Estimation—A Comparison, Mechanical Systems and Signal Processing, vol. 11, pp. 791-809, Nov. 1997.

[88]. Kamarti, S.V., Kumara, S.R.T., P.H. Cohen, Flank wear estimation in turning through wavelet representation of acoustic emission signals, Journal of Manufacturing Science and Engineering, Transactions of the ASME 122 (2000) 12–19.

[89]. Kassim, A. A., Zhu M, and Mannan, M. A., Tool condition classification using Hidden Markov Model based on fractal analysis of machined surface textures, Machine Vision and Applications (2006) 17:327–336

[90]. Khraisheh, M. K., Pezeshki, C. and Bayoumi, A. E., Time Series Based Analysis for Primary Chatter in Metal Cutting, Journal of Sound and Vibration, 1995, 180(1), 67-87.

[91]. Kim, J.D., and Kim, D.S., Theoretical analysis of micro-cutting characteristics in ultra-precision machining, Journal of materials Processing Technology 49 (1995) 387–398.

[92]. King, R. I. Ed., Handbook of High-Speed Machining Technology, Chapman and Hall: New York, 1985

[93]. Kirby. M., Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns, Wiley-Interscience, 2000

[94]. Klocke F., Reuber M., and Kratz H., Application of a wavelet-based signal analysis for evaluating the tool state in cutting operations, Industrial Electronics Society, 2000. IECON 2000. 26th Annual Conference of the IEEE Volume 3, 22-28 Oct. 2000 Page(s):1967 - 1972 vol.3.

[95]. Konig, W., Kutzner, K., and Schehl, U., Tool monitoring of small drills with acoustic emission, International Journal of Machine Tools and Manufacturing 32 (1992) 487493.

[96]. Kothamasu, R., Huang, S. H., and Verduin, W. H., Comparison of Computational Intelligence and Statistical Methods In Condition Monitoring for Hard Turning, International Journal of Production Research, Vol. 43, No. 3, 1 February 2005, 597–610

[97]. Kumar, S.A., Ravindra, H. V. and Srinivasa, Y.G. In-process Tool Wear Monitoring through Time Series Modeling and Pattern Recognition, Int. J. Prod. Res., Vol. 35, No. 3, pp. 739-751. 1997.

[98]. Kurada, S. and C. Bradley, A review of machine vision sensors for tool condition monitoring. Computers in Industry. 34(1): p. 55-72, 1997.

[99]. Kwak, J.S. and Ha, M.K., Detection of dressing time using the grinding force signal based on the discrete wavelet decomposition, The International Journal of Advanced Manufacturing Technology, Volume 23, 2004, Numbers 1-2 pp 87 - 92.

[100]. Kwak, J.S. Application of wavelet transform technique to detect tool failure in turning operations, The International Journal of Advanced Manufacturing Technology, May 2006, Volume 28, Numbers 11-12, pp.1078 - 1083

[101]. Cohen, L., Time–frequency distribution—A review, Proceedings of the IEEE 77 (1989) 941–981.

[102]. Lee, B.Y., Tarng, Y, S, Application of the discrete wavelet transform to the monitoring of tool failure in end milling using the spindle motor current. International Journal of Advanced Manufacturing Technology, 15:238–243, (1999).

[103]. Lee, J.H., Kim, D.E., and Lee, S.J., Application of neural networks to flank wear prediction, Mechanical Systems & Signal Processing, vol. 10 (3), pp. 265-276, 1996

[104]. Lee, S. K. and White, P. R., Higher-Order Time–Frequency Analysis And its Application to Fault Detection In Rotating Machinery, Mechanical Systems and Signal Processing 1997, Vol. 11(4), pp.637-650

[105]. Leem C.S., Dornfeld D.A., Design and Implementation of Sensor based Tool wear Monitoring Systems, Mechanical Systems and Signal Processing, Volume 10, Number 4, July 1996, pp. 439-458.

[106]. Lewicki M.S., and Sejnowski, T. J., Learning Overcomplete Representations. Neural computation, 12(2): 337–365, 2000.

[107]. Li X., Yao Y. and Yuan Z., On-line tool condition monitoring system with wavelet fuzzy neural network, Journal of Intelligent Manufacturing (1997) 8, 271-276

[108]. Li, D., and Douglas, O. S., Speech processing: A Dynamic and Object-oriented Approach, Marcel Dekker, Inc, 2003.

[109]. Li, S., and Elbestawi, M.A., Fuzzy Clustering for Automated Tool Conditon Monitoring in Machining, Mechanical Systems and Signal Processing 1996 10(5) 533-550

[110]. Li, W., Gu, F., Ball, A. D., Leung, A. Y. T. and Phipps, C. E., A study of the noise from diesel engines using the independent component analysis, Mech. Systems & Signal Processing Vol.15(6), pp.1165-1184,2001

[111]. Li, X., and Guan, X.P., Time-frequency-analysis-based minor cutting edge fracture detection during end milling, Mechanical Systems and Signal Processing 18 (2004) 1485–1496.

[112]. Li, X., (1999) On-line detection of the breakage of small diameter drills using current signature wavelet transform. International Journal of Machine Tools & Manufacture, 39:157–164

[113]. Li, X., Detection of tool flute fracture in end milling using feed-motor current signatures, IEEE/ASME Transactions on Mechatronics 6 (2001) 491–498.

[114]. Li, X., Dong, S., and Yuan, Z., Discrete wavelet transform for tool breakage monitoring, International Journal of Machine Tools & Manufacture 39 (1999) 1935–1944.

[115]. Li, X., Tso, S. K., and Jun Wang (2000): Real time tool condition monitoring using wavelet transforms & fuzzy techniques. IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, Vol. 31, No. 3, pp. 352-357, 2000.

[116]. Li, X., Yuan, Z. Tool wear monitoring with wavelet packet transform-fuzzy clustering method, Wear. Vol. 219, No. 2, 1998, pp. 145-154.

[117]. Liang S. Y., Rogelio L. H., Robert G. L., Machining Process Monitoring and Control: The State-of-the-Art, Journal of Manufacturing Science and Engineering, MAY 2004, Vol. 126, pp297-310.

[118]. Liang, S.Y. and Dornfeld, D.A. Detection of Cutting Tool-wear Using Adaptive Time Series Modeling of Acoustic Emission Signal, Sensors for Manufacturing, ASME, pp. 27-38. 1987.

[119]. Lin, J., Zuo, M.J., and Fyfe, K.R., Mechanical fault detection based on the wavelet de-noising technique, Journal of Vibration and Acoustics 126 (2004) 9–16.

[120]. Lin, J.and Qu, L., Journal of Sound and Vibration, Feature Extraction Based On Morlet Wavelet and Its Application for Mechanical Fault Diagnosis, Journal of Sound and vibration (2000) 234(1), 135-148

[121]. Lin, S.C., and Lin, R.J., Tool wear monitoring in face milling using force signals, Wear 198 (1996) 136-142

[122]. Liu, B., and Ling, S. F, On the Selection of Informative Wavelets for Machinery Diagnosis, Mechanical Systems and Signal Processing, 13(1), 145-162, 1999.

[123]. Liu, B., Selection Of Wavelet Packet Basis For Rotating Machinery Fault Diagnosis, Journal Of Sound And Vibration 284 (2005) 567–582

[124]. Liu, K., Process Modeling of Micro-Cutting Including Strain Gradient Effects, PhD Thesis, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, December, 2005.

[125]. Liu, M., Liang, S.Y., Analytical modeling of acoustic emission for monitoring of peripheral milling process, Int. J. Mach. Tool Manufacture., 31 (4) (1991) 589-606

[126]. Liu, X., Devor, R. E., Kapoor, S. G., and Ehmann,K. F., The Mechanics of Machining at the Micro-Scale: Assessment of the Current State of the Science, Journal of Manufacturing Science and Engineering, 126:4, 666-677, 2004

[127]. Loutridis, S., and Trochidis, A., Classification Of Gear Faults Using Hoelder Exponents, Mechanical Systems And Signal Processing 18 (2004) 1009–1030.

[128]. Luis A F.G., Gilberto H.R., Rocio P.V, Rene de Jesus R.T. and Wbaldo L.T, Sensorless tool failure monitoring system for drilling machines, International Journal of Machine Tools & Manufacture 46 (2006) 381–386

[129]. MacKay, D. J., 1992, A practical Bayesian framework for back-propagation networks. Neural Computation, 4(3), 448–472.

[130]. Malhi, A., and Gao, R.X., PCA-Based Feature Selection Scheme for Machine Defect Classification, IEEE Transactions on Instrumentation and Measurement, Vol.53, No. 6, December 2004.

[131]. Mallat, S. G. A theory of multiresolution signal decomposition: The wavelet representation. IEEE Trans. on Pattern and Machine Intelligence, 11(7):674-693 1989.

[132]. Mallat, S. G. A Wavelet Tour of Signal Processing, San Diego: Academic Press, 1998.

[133]. Mallat, S. G. and Hwang, W. L., Singularity detection and processing with wavelets. IEEE Transactions on Information Theory 38, 617–643, 1992.

[134]. Mallat, S. G. and Zhang, Z., Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 41, 1993, pp 3397-3415.

[135]. Mallat, S. G., Krim, H., Tucker, D., Donoho. D., On denoising and best signal representation, IEEE Transactions on Information Theory, Vol. 45, No. 7, 2225-2238, November, 1999

[136]. Mallat, S., and Zhong, S., Characterization of Signals from Multiscale Edges, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No.7, July 1992.

[137]. Mao, K. Z., Identifying Critical Variables of Principal Components for Unsupervised Feature Selection, IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 35, No. 2, pp339-344, April 2005

[138]. Menon, S., Schoess, J.N., Hamza, R., and Busch, D., Wavelet-based acoustic emission detection method with adaptive thresholding, Proceedings of SPIE 3986 (2000) 71–77.

[139]. Mori K, Kasashima N, Fu JC, and Muto K (1999) Prediction of small drill bit breakage by wavelet transforms and linear discriminant functions. International Journal of Machine Tools & Manufacture, 39:1471–1484.

[140]. Muller, P. and Vidakovic, B. (1998), Bayesian inference with wavelets: Density estimation, J. Computational and Graphical Statistics, 7(4), 456–468.

[141]. Nason, G.P. Wavelet shrinkage using cross-validation. J. R. Statist. Soc. B, 58 (1996), 463–479.

[142]. Neal, R. M. (1996). Bayesian learning for neural networks. No. 118 in Lecture Notes in Statistics. New York: Springer.

[143]. Niu, Y.M., PhD thesis, Information-driven tool condition monitoring techniques, Dept. of Mechanical Engineering, Faculty of Engineering, National University of Singapore, 2001.

[144]. Niu, Y.M., Wong, Y.S. and Hong, G.S., Multi-category classification of tool conditions using wavelet packets and ART2 network, ASME J. of Manufacturing Science and Technology, Vol. 120, 807-815, 1998.

[145]. Obikawa, T., and Shinozuka, J., Monitoring of flank wear of coated tools in high speed machining with a neural network ART2, International Journal of Machine Tools & Manufacture 44 (2004) 1311–1318

[146]. Ogden, R. T. and Lynch, J. D. (1999) Bayesian analysis of change-point models. Lecture Notes Statistics, 141, pp 67-82.

[147]. Oguamanam D.C.D., Raafat H. and Taboun S.M. A Machine Vision System for Wear Monitoring and Breakage Detection of Single-point Cutting Tools, Computers in Industrial Engineering, Vol. 26, No. 3, pp. 575-598. 1994.

[148]. Oppenheim, A. V., Willsky, A. S., Nawab. S. N., Nawab. S., Signals and Systems, 2nd Edition, Prentice Hall, 1996.

[149]. Owsley L.M.D., and Atlas L. E., Bernard G. D., Self-organizing feature maps and hidden Markov models for machine-tool monitoring, IEEE Transactions on Signal Processing, v 45, n 11, Nov, 1997, p 2787-2798.

[150]. Pai, P.S., and Rao, P.K.R., Acoustic emission analysis for tool wear monitoring in face milling, Int. J. Prod. Res., 2002, vol. 40, no. 5, 1081-1093

[151]. Pasek, Z. J., Exploration of rough sets theory use for manufacturing process monitoring, IMechE 2006 Proc. IMechE Vol. 220 Part B: J. Engineering Manufacture, pp335-374

[152]. Peng, Z., He, Y., Chen Z., and Chu, F., Identification of the shaft orbit for rotating machines using wavelet modulus maxima, 2002 Mechanical Systems and Signal Processing 16, 623–635.

[153]. Peng, Z.K., and Chu, F.L., Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing 18 (2004) 199–221.

[154]. Peng, Z.K., Chu, F.L., and Tse, P. W., Singularity Analysis of The Vibration Signals By Means Of Wavelet Modulus Maximal Method, Mechanical Systems And Signal Processing In Press.

[155]. Peng, Z.K., Tse, P. W., and Chu, F.L., An improved Hilbert–Huang transform and its application in vibration signal analysis, Journal of Sound and Vibration 286 (2005) 187–205.

[156]. Penny, W.D., and Roberts, S.J., Bayesian neural networks for classification: how useful is the evidence framework? Neural Networks 12 (1999) 877–892.

[157]. Percival, D.B. and Walden, A.T. Wavelet Methods for Time Series Analysis. Cambridge: Cambridge University Press, 2000.

[158]. Pittner, S., and Kamarthi, S.V., Feature Extraction from Wavelet Coefficients for Pattern Recognition Tasks, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.1, January 1999.

[159]. Qiu, H., Lee, J., and Lin, J., and Yu, G., Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration 289 (2006) 1066–1090.

[160]. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[161]. Rahman, R., Kumar, A. S., Prakash, J. R. R., Micro-milling of pure copper, International Journal of Materials Processing Vol. 16 pp.39-43, 2001

[162]. Rivet, B., Vigneron, V., Anisoara P.I, and Jutten, C., Wavelet de-noising for blind source separation in noisy mixtures, The fifth internaltional conference ICA 2004, Granada, Spain, LNCS 3195, pp 263-270, 2004.

[163]. Roan, M. J., Erling, J. G. and Sibul, L. H., A new,non-linear, adaptive, blind source separation approach to geartooth failure detection and analysis, Mechanical Systems & Signal Processing Vol.16(5), pp.719–740,2002

[164]. Roberts, S. and Everson, R. . Independent Component Analysis: principles and practice. Cambridge University Press, March 2001.

[165]. Robertson, A. N., Farrar C. R., and Sohn, H., Singularity Detection For Structural Health Monitoring Using Holder Exponents, Mechanical Systems And Signal Processing (2003) 17(6), 1163–1184.

[166]. Ruggeri, F. and Vidakovic, B., A Bayesian decision theoretic approach to the choice of thresholding parameter, Statistica Sinica 9 (1999), 183-197

[167]. Safavi, A.A., Chen, J., and Romagnoli, J.A., Wavelet-based Density Estimation and Application to Process Monitoring, Artificial Intelligence in Chemical Engineering, May 1997, Vol. 43, No.5, pp1227-1241.

[168]. Saglam, H., and Unuvar, A., Tool Condition Monitoring in Milling on cutting forces by a Neural Network. Int. J. Production Research, 2003, Vol. 41, No. 7, 1519–1532

[169]. Saito N., Coifman, R. R., Geshwind, F. B., and Warner, F., Discriminant feature extraction using empirical probability density estimation and a local basis library, Pattern Recognition, vol. 35, pp.2841-2852, 2002

[170]. Scheffer C. and Heyns P.S, Wear monitoring in turning operations using vibration and strain measurements, Mechanical Systems and Signal Processing, v 15, n 6, Nov. 2001, p 1185-202

[171]. Scheffer, C., Engelbrecht,H., and Heyns, P. S., A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear, Neural Computation and Applications (2005) 14: 325-336

[172]. Serviere, C., and Fabry, P., Blind source separation of noisy harmonic signals for rotating machine diagnosis, Journal of Sound and Vibration (in press)

[173]. Shaw, M.C., Metal cutting principles, 2nd ed , New York : Oxford University Press, 2005

[174]. Sick, B., Review: 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 (2002) 16(4), 487–546.

[175]. Silva, R. G., Reuben, R. L., Baker, K. J. and Wilcox, S. J., Tool Wear Monitoring Of Turning Operations By Neural Network And Expert System Classification Of A Feature Set Generated From Multiple Sensors, Mechanical Systems and Signal Processing

[176]. Silverman, B.W., Density estimation for statistics and data analysis, New York : Chapman and Hall, 1986

[177]. Staszewski W.J., Worden K., and Tomlinson G.R. Time-frequency analysis in gearbox fault detection using the Wigner-Ville distribution and pattern recognition, Mechanical Systems and Signal Processing, v 11, n 5, Sept. 1997, p 673-92

[178]. Staszewski, W. J., Identification of Non-Linear Systems Using Multi-Scale Ridges and Skeletons of The Wavelet Transform, Journal of Sound and Vibration 1998 214 (4), pp.639-658.

[179]. Suh, C.S., Khurjekar, P.P., and Yang, B., Characterisation And Identification Of Dynamic Instability Inmilling Operation, Mechanical Systems And Signal Processing (2002) 16(5), 853–872

[180]. Sun, J., Hong, G. S., Rahman, M., and Wong, Y. S., Identification of feature set for effective tool condition monitoring by acoustic emission sensing, International Journal of Production Research, 2004, vol. 42, no. 5, 901-918

[181]. Sun, J., Hong, G.S., Wong, Y.S., Rahman, M., and Wang, Z.G., Effective training data selection in tool condition monitoring system, International Journal of Machine Tools & Manufacture 46 (2006) 218–224.

[182]. Sun, Q., Tang; Y., Lu, W., Y.,.and Ji, Y., Feature extraction with discrete wavelet transform for drill wear monitoring, Journal of Vibration and Control, v 11, n 11, Nov. 2005, p 1375-96

[183]. Sutter, G., and Molinari, A., Analysis of the Cutting Force Components and Friction in High Speed Machining, pp 245-250 Vol. 127, MAY 2005, ASME, Journal of Manufacturing Science and Engineering.

[184]. Tansel, I.N., Bao, W.Y., Reen, N.S., and Kropas-Hughes, C.V., Genetic tool monitor (GTM) for micro-end-milling operations, International Journal of Machine Tools & Manufacture 45 (2005) 293–299

[185]. Tansel, I.N., Arkan, T.T., Bao, W.Y., Mahendrakar, N., Shisler, B., Smith, D., and M. McCool, Tool wear estimation in micro-machining. Part I: tool usage–cutting force relationship, Int. J. Mach. Tools Manufacture Vol. 40 pp.599–608, 2000

[186]. Tansel, I.N., Mekdeci, C., and McLaughlin, C., Detection of Tool Failure in End Milling with Wavelet Transformations and Neural Networks (WT-NN), International Journal of Machine Tools & Manufacture, vol. 35, pp. 1137-1147, Aug. 1995.

[187]. Tansel, I.N., Mekdeci, C., Rodriguez, O. and Uragun, B. Monitoring Drill Conditions with Wavelet Based Encoding and Neural Network. International Journal of Machine Tools & Manufacture, 1993, Vol. 33, No. 4, pp. 559–575.

[188]. Tansel, I.N., Rodriguez, O., Trujillo, M., E. Paz, Li, W., Micro-end-milling—III. wear estimation and tool breakage detection using acoustic emission signals, International Journal of Machine Tools & Manufacture 38 (1998) 1449-1466

[189]. Tansel, I.N., Rodriguez, O., Trujillo, M., Paz, E., and Li, W., Micro-end-milling—I. wear and breakage, International Journal of Machine Tools & Manufacture Vol. 38 pp.1419-1436, 1998.

[190]. Tarng, YS, and Lee, BY, A sensor for the Detection of Tool Breakage in NC Milling, Journal of Materials Processing Technology, Vol.36, pp.259-272.1993

[191]. The FastICA homepage. http://www.cis.hut.fi/projects/ica/fastica/

[192]. Theodoridis, S. and Koutroumbas, K., Pattern Recognition, Academic Press, 2003

[193]. Tian X., Lin, J., Fyfe, K.R. and Zuo, ,M. J. Gearbox fault diagnosis using independent component analysis in the frequency domain and wavelet filtering, IEEE ICASSP 2003 II245-248.

[194]. Tonshoff, H.K., Li, X., and Lapp, C., Application of fast Haar transform and concurrent learning to tool-breakage detection in milling, IEEE/ASME Transactions on Mechatronics, 8 (2003) 414–417.

[195]. Trujillo, M., Li, W., Fallerio, B., Paz, E., and Tansel, I.N., Inspection of Micro-tools at high rotational speeds, International Journal of Machine Tools & Manufacture 34 (1994) 1059–1077

[196]. Tse, P. W., Yang, W. X., and Tam, H.Y., Machine Fault Diagnosis Through an Effective Exact Wavelet Analysis, Journal of Sound And Vibration 277 (2004) 1005–1024

[197]. Tumer, I. Y.,Wood, K. L., and Busch-Vishniac, I. J., Monitoring Of Signals From Manufacturing Processes Using The Karhunen-LoeeVe Transform, Mechanical Systems And Signal Processing (2000) 14(6), 1011-1026.

[198]. Tumer, I., Wood, K., and Busch-Vishniac, I. 1997. Monitoring fault condition during manufacturing using the Karhunen-Loeve transform. In 1997 ASME Mechanical Vibration and Noise Conference, System Health Monitoring Symposium, volume DETC97-VIB4234.

[199]. Vannucci, M. (1995). Nonparametric Density Estimation using Wavelets: A Review. Discussion Paper 95-26, ISDS, Duke University, USA.

[200]. Vidakovic, B., Statistical modeling by wavelet, John Wiley & Sons, 1999.

[201]. Vogler, M. P., Devor, R. E. and Kapoor, S. G. On the Modeling and Analysis of Machining Performance in Micro-End milling, Part II: Cutting Force Prediction, Journal of Manufacturing Science and Engineering, 126:4, 694-704, 2004

[202]. Wang, C., and Gao, R.X., Wavelet transform with spectral post-processing for enhanced feature extraction, IEEE Transactions on Instrumentation and Measurement 52 (2003) 1296–1301.

[203]. Wang, L., Mehrabi, M. G., and Elijah K.A., Hidden Markov Model-based Tool Wear Monitoring in Turning, ASME Journal of Manufacturing Science and Engineering, 652 -658, Vol. 124, August 2002.

[204]. Wang, W. H., Hong, G. S., Wong Y. S., and Zhu, K. P., Sensor fusion for on-line tool condition monitoring in milling, International Journal of Production Research, In Press.

[205]. Wang, W., Jones, P., Partridge,D., A Comparative Study of Feature-Salience Ranking Techniques, Neural Computation 13, 1603–1623 (2001)

[206]. Weck, M., Fischer, S., and Vos, M., 1997, Fabrication of Microcomponents Using an Ultraprecision Machine Tool, Nanotechnology, 8 (3), pp. 145–148.

[207]. Weule, H., Huntrup, V., and Tritschle, H., 2001, Micro-Cutting of Steel to Meet New Requirements in Miniaturization, CIRP Ann., 50, pp. 61–64.

[208]. Wickerhauser, M.V. and Coifman,R.R., Entropy based methods for best basis selection. IEEE Trans. Inf. Th., 38(2):719-746, 1992.

[209]. WTEC Panel Report on International Assessment of Research and Development in Micromanufacturing Final Report, World Technology Evaluation Center (WTEC), Inc. Baltimore, Maryland, USA. October 2005.

[210]. Wu, Y. and Du, R., Feature extraction and assessment using wavelet packets for monitoring of machining process, Mechanical Systems and Signal Processing 1996 10(1) 29-53.

[211]. Wu, Y., P. Escande, and Du, R., A New Method for Real-Time Tool Condition Monitoring in Transfer Machining Stations, ASME Journal of Manufacturing Science and Engineering, MAY 2001, Vol. 123, pp 339-347.

[212]. Yen, G. G., and Lin, K. C., Wavelet Packet Feature Extraction for Vibration Monitoring, IEEE Transitions on Industrial Electronics, Vol. 47, No. 3, June 2000.

[213]. Yesilyurt. I., End mill breakage detection using mean frequency analysis of scalogram, International Journal of Machine Tools & Manufacture, In Press.

[214]. Yoon, M C and Chin, D H , Cutting force monitoring in the end milling operation for chatter detection, IMechE 2005 Proc. IMechE Vol. 219 Part B: J. Engineering Manufacture, pp455-465

[215]. Zhang, S., Mathew, J., Ma, L., and Sun, Y., Best Basis-based in Machine Fault Diagnosis, Mechanical Systems and Signal Processing Vol. 19, Issue 2, March 2005, Pages 357-370.

[216]. Zhou, Q., Hong, G. S., and Rahman, M., A New Tool Life Criterion for Tool Condition Monitoring Using a Neural Network, Engineering Applications of Artificial Intelligence, Vol. 8, No. 5, pp. 579-588, 1995.

[217]. Zhu, K.P., Hong, G. S., Wong, Y. S., and Wang, W.H., Cutting Force Denoising in Micro-milling Tool Condition Monitoring, International Journal of Production Research, In Press, 2006.

[218]. Zhu K.P., Hong G.S., Wong Y.S., Sequential ICA for cutting force denoising in micromachining tool wear monitoring, ICMA 2004, October, 2004, Wuhan, China

[219]. Zhu, R., R. E. Devor, and Kapoor, S. G., A Model-Based Monitoring and Fault Diagnosis Methodology for Free-Form Surface Machining Process, ASME Journal of Manufacturing Science and Engineering, August 2003, Vol. 125

[220]. Zibulevsky, M. and Pearlmutter, B.A. Blind Source Separation by Sparse Decomposition, Neural Computations 13(4), 2001.

Appendix A

A.1 Flank wear-force correlation

Correlation between the cutting forces and tool flank wear:

$$\rho = \left| \frac{\sum_i (q_i - \bar{q})(V_i - \bar{V})}{\sum_i (q_i - \bar{q})^2 \sum_i (V_i - \bar{V})^2} \right| \times 100\% \qquad \text{(A.1)}$$

where q_i and V_i are the force component and corresponding wear.

The correlations were calculated as following:

Fx: $\rho_{x1} = 73\%$ (mean force) $\quad \rho_{x2} = 71\%$ (dynamic component)

Fy: $\rho_{y1} = 89\%$ (mean force) $\quad \rho_{y2} = 86\%$ (dynamic component)

Fz: $\rho_{z1} = 54\%$ (mean force) $\quad \rho_{z2} = 63\%$ (dynamic component)

A.2 Linear Regression Analysis

The experiments have shown that the mean force Fy is mostly correlated with tool flank wear. For the linear regression (figure A.1), it is shown that the linear regression can not fit well, the variance is high.

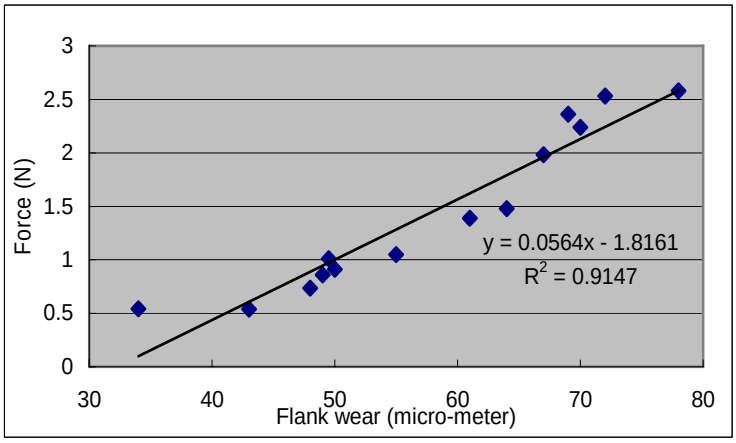


Figure A.1 Flank wear linear regression with Fy

A. 3 MLP Analysis

The MLP mapping of flank wear-force relationship was done with two similar working conditions. One is used to train the MLP and the other is used for testing the MLP.

The tool wear-forces pairs were trained with a 4-4-1 three-layer MLP. Inputs were chosen for the most correlated four force components, mean force of Fx, Fy and dynamic Component: Fx, Fy. The neural network output is the estimated flank wear. The trained MLP network is used for testing the other test. However, the testing (figure 4.21) shows that it is not a good generalization, the error is always large and the maximum error reaches 31.2%.

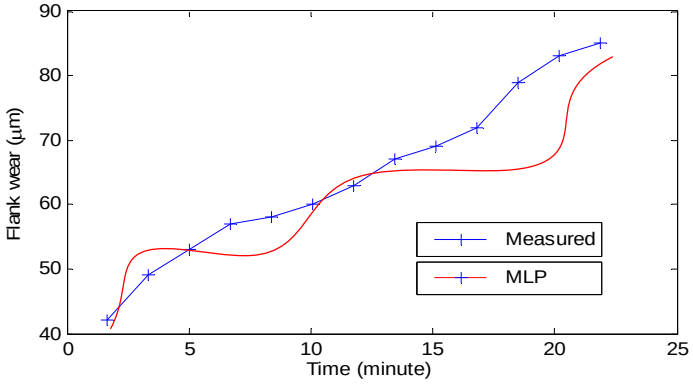


Figure A.2 the testing result with MLP

A.4 Result

Both linear regression and MLP neural networks were used to estimate flank wear for milling processes under various cutting conditions. However, the estimation of flank wear shows that there is no good generalization for both linear regression and MLP.

Appendix B:

Multi-resolution analysis (MRA)

Continuous wavelet transforms (CWTs) are recognized as effective tools for both stationary and non-stationary signals. However, it involves a lot of redundant information and very slow in computation. Fast Wavelet Transform (FWT) was developed by Mallat with discrete wavelet transform (DWT). It was based on the idea of multiresolution (MRA) and constructed with *Conjugate Quadratic Filters* (CQF) [131].

Both the construction and applications of orthonormal wavelets use the important concept of multi-resolution analysis developed by Mallat (1989), [131], [132]. A multi-resolution analysis consists of a family of closed subspaces $\{V_j\}_{j\in Z}$ of $L^2(R)$ satisfying

1) $\forall(j,k)\in Z^2,\ \ f(t)\in V_j \Leftrightarrow f(t-2^j k)\in V_j$ (B.1)

2) $\forall j\in Z,\ \ V_j \supset V_{j+1},\ \cdots V_0 \supset V_1 \supset V_2 \cdots V_j \supset V_{j+1}\cdots$ (B.2)

3) $\forall j\in Z,\ f(t)\in V_j \Leftrightarrow x(t/2)\in V_{j+1}$ (B.3)

4) $\underset{j\to\infty}{Lim} V_j = \bigcap_{j=-\infty}^{\infty} V_j = \{0\}$ (B.4)

5) $\underset{j\to\infty}{Lim} V_j = Closure(\bigcup_{j=-\infty}^{\infty} V_j) = L^2(R)$ (B.5)

6) There exists θ such that $\{\theta(t-n)\}_{n\in Z}$ is a Riesz basis for V_0 (B.6)

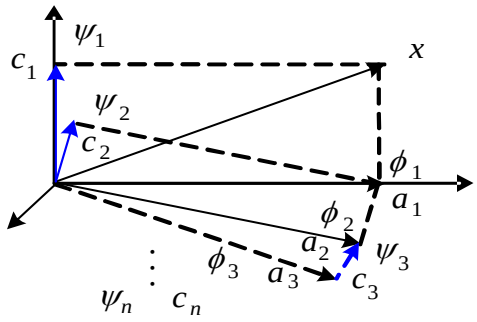
This is Mallat algorithm. If the coefficients of two scale equation is looked as filter, then Mallat algorithm is in reality two-channel filter banks. In the sense, scale function and wavelet are known as low-pass filter and high-pass filter. By defining W_j as an orthogonal complement of V_j in V_{j-1}, i.e. $V_{j-1} = V_j \oplus W_j$ and $V_j \perp W_j$.

The space W_j are the differences between the V_j. That is W_j contains the details and V_j contains coarse (approximation) information of f at level j.

The space V_j are the sums of W_j,

$$V_0 = W_1 \oplus W_2 \oplus ... \oplus W_j \oplus W_{j+1} \oplus V_{j+1} \qquad \text{(B.7)}$$

This geometric construction can be viewed through figure A.1.



$$\begin{aligned} x &= c_1\psi_1 + a_1\phi_1 \\ &= c_1\psi_1 + c_2\psi_2 + a_2\phi_2 \\ &= c_1\psi_1 + c_2\psi_2 + c_3\psi_3 + a_3\phi_3 \\ &= c_1\psi_1 + c_2\psi_2 + c_3\psi_3 + ... + c_n\psi_n + a_n\phi_n \end{aligned}$$

Figure A.1 Geometry of MRA

Appendix C:

Three Problems of Hidden Markov Models and their solutions

B.1 Three Fundamental Problems for HMM

There are three problems of interest that must be solved for the model:

Problem 1:Evaluation

Given the observation sequence $O=o_1\ o_2\ \ldots o_T$ and a model λ, compute $P(O|\ \lambda)$, the probability of the observation sequence, given the model. This problem is evaluation or scoring problem. If we consider the case in which we are trying to choose among several models, this solution give us the model which best matches the observation.

Problem 2: Decoding

Given the observation sequences and the model, find the optimal corresponding state sequence. This is the one that tries to uncover the hidden part of the model. There is no exact and unique solution for this problem, but in practice, an optimality criterion is considered to solve the problem. There are several optimality criteria that can be applied.

Problem 3: Training

Given some observations sequences, how to estimate model parameters.

B.2 Evaluation Problem

The most straightforward way to find the probability of observation given the model, is enumerating every possible state sequence of number of observations T. The probability of the observation sequence for the sate sequence of $Q=\{q_1,q_2,...q_t\}$ is

$$P(O,Q\mid\lambda)=b_{q1}(o_1)b_{q2}(o_2)....b_{qT}(o_T)$$

and the probability of such a state sequence can be written

$$P(Q \mid \lambda) = \pi_{q1} a_{q1q2} a_{q2q3} .... a_{qT-1qT}$$

So, the probability of observation given the model will be

$$P(O \mid \lambda) = \sum_{all\ Q} P(Q \mid \lambda) P(O, Q \mid \lambda)$$
$$= \sum_{all\ Q} \pi_{q1} b_{q1}(o_1) a_{q1q2} b_{q2}(o_2) .... a_{qT-1qT} b_{qT}(o_T)$$

Direct calculation of this equation will involve on he order of $2TN^T$ calculations that is absolutely impossible for practical applications. Fortunately an efficient procedure exists and is called forward-backward procedure.

Forward-Backward Algorithm

Consider forward variable $\alpha_t(i)$ defined as

$$\alpha_t(i) = P(o_1 o_2 ... o_T, q_t = S_i \mid \lambda)$$

Here is the procedure to compute this variable inductively:

- Initialization

$$\alpha_1(i) = b_j(o_1) \quad 1 \le i \le N$$

- Induction

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad 1 \le t \le T-1$$

- Termination

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

We see that it requires on the order of N^2T calculations. In similar manner, we can define backward variable as follows:

$$\beta_t(i) = P(o_{t+1} o_{t+2} ... o_T \mid q_t = S_i, \lambda)$$

Again, we can solve for this variable inductively,

- Initialization $\beta_T(i) = 1 \quad 1 \le i \le N$

- Induction $\beta_t(i) = \sum_{i=1}^{N} b_j(o_{t+1})\beta_{t+1}(j)a_{ij} \quad 1 \le t \le T-1$

Backward procedure will be used in the solution to problem 3and it is not required for the solution of problem 1.

B. 3 Decoding problem

There are several possible was of solving problem 2, the optimal state sequence associated with given observation. For example one optimality criterion is to choose the states which are individually most likely. To implement this, define a new variable:

$$\gamma_t(i) = P(q_t = S_i) = \frac{\alpha_t(i)\beta_t(i)}{P(O \mid \lambda)}$$

Using this variable, we can solve for individually most likely state at time *t*:

$$q_t = \arg\max\left[\gamma_t(i)\right] \quad 1 \le i \le N$$

But this solution is not perfect solution in case of there is some null transition between the states and this solution determines the most likely state without regard to the probability of occurrence of sequences of states. So, we need to modify the optimality criterion. Following algorithm find the single best state sequence for the given observation sequence and the model.

The best score along the a single path at time t, which accounts for the first *t* observations and ends in state S_i can be expressed as follows:

$$\delta_t(i) = \max P(q_1 q_2 ... q_t = i, o_1 o_2 ... o_t \mid \lambda)$$

The complete procedure for finding the best state sequences follows: (ψ is the variable that track the argument which maximized)

Viterbi Algorithm

Initialization:

$$\delta_1(i) = \pi_i b_i(o_1) \qquad 1 \le i \le N$$
$$\psi_1(i) = 0$$

Recursion

$$\delta_t(i) = \max\left[\delta_{t-1}(i)a_{ij}\right]b_j(o_t) \quad 2 \le t \le T$$

$$\psi_t(i) = \arg\max\left[\delta_{t-1}(i)a_{ij}\right] \quad 1 \le i \le N$$

Termination

$$P^* = \max\left[\delta_T(i)\right] \quad 2 \le t \le T$$

$$q_T{}^* = \arg\max\left[\delta_T\right] \quad 1 \le i \le N$$

Path backtracking

$$q_T{}^* = \psi_{t+1}(q_{t+1}{}^*) \quad t = T-1, T-2, \ldots 1$$

B.4 Training problem

There is no known analytical approach to solve the model parameters that maximizes probability of observation given that model. Here one of the most famous algorithms named the expectation-modification algorithm is described.

For this algorithm, again, a new variable is defined:

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(i)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(i)}$$

The re-estimation procedure here is as follows:

$$\pi_i = \gamma_1(i),\ a_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)},\ b_j(k) = \frac{\sum_{t=1,O_t=V_k}^{T}\gamma_t(i)}{\sum_{t=1}^{T}\gamma_t(i)}$$

This procedure will be repeated until convergence of model parameters. This formulation is for single discrete observation sequence. As it was explained before we have continuous observation in most of the real-world applications. In addition to this matter, for appropriate training of the model, we need to feed multi-observation sequences to the re-estimation procedure. The modification for the re-estimation procedure is straightforward: suppose we have the set of *K* observation of sequences. Therefore we need to maximize the product of each probability of individual observation given the model instead of the one we saw before.

$$P(\underline{O} \mid \lambda) = \prod_{k=1}^{K} P(o^{(k)} \mid \lambda) = \prod_{k=1}^{K} P_k \quad that \quad \underline{o} = \left[o^{(1)} o^{(2)} ... o^{(K)} \right]$$

All of the parameters used for intermediate computation including forward variable and backward variables will be computed individually for each observation; $\alpha_t^{(k)}(i), \beta_t^{(k)}(i), \gamma_t^{(k)}(i)$.

The final reestimation formulation for ergodic, continuous observation HMM with multi-observation training can be shown this way. The term "ergodic" here refers to this fact that every state the model could be reached in a single step from every other state (fully connected HMM).

This formulation is supposed for mixture of Gaussian distribution as *pdf* of observations.

$$\pi_i = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \gamma_1^{(k)}(i)}{\sum_{k=1}^{K} \frac{1}{P_k}}, \; a_{ij} = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \alpha_t^{(k)}(i) a_{ij} b_j(o_{t+1}^{(k)}) \beta_{t+1}^{(k)}(j)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \alpha_t^{(k)}(i) \beta_t^{(k)}(j)}$$

$$\mu_i = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \gamma_t^{(k)}(j).o_t^{(k)}}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \gamma_t^{(k)}(j)}, \; \sigma_i^2 = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \gamma_t^{(k)}(j).(o_t^{(k)} - \mu_i)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T} \gamma_t^{(k)}(j)}$$

In this process forward and backward variables consist of a large number of terms of a and b that are generally significantly less than 1. It can be seen as t get big, each term will exceed precision range of any machine even in double precision. Hence the only reasonable way of those computations is incorporating a scaling procedure.

Publications

1. Accepted

Zhu K.P., Hong G.S., Wong Y.S., Sequential independent component analysis for cutting force denoising in micromachining tool wear monitoring, *International Conference of Manufacturing Automation (ICMA) 2004*, October, 2004, Wuhan, China.

Zhu K.P., Hong G. S., Wong Y. S. Wang W.H., Cutting force denoising in micro-milling tool condition monitoring, *International Journal of Production Research,* In Press.

Wang W. H., Hong G. S. Wong Y. S., and Zhu K.P., Sensor fusion for on-line tool condition monitoring in milling, *Journal of Production Research,* In Press.

Zhu K.P., Wong Y. S., Hong G. S., Noise-robust tool condition monitoring in micro-milling with hidden Markov models, book chapter in *Soft Computing Applications in Industry*, Springer-Verlag, 2007.

2. Submitted

Zhu K.P., Hong G. S., Wong Y. S., Discriminate feature selection for hidden Markov models in micro-milling tool wear classification, submitted to *Machining Science and Engineering.*

Zhu K.P., Wong Y. S., Hong G. S., Multi-category micro-milling tool wear classification with continuous hidden Markov models, submitted to *Mechanical System and Signal Processing*

Zhu K.P., Wong Y. S., Hong G. S., Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results. Submit to *ASME transaction on Manufacturing Science and Engineering*.

3. In Progress

Zhu K.P., Hong G. S., Wong Y. S., Multiscale analysis of cutting forces for in micro milling tool condition monitoring, under revision.