Int J Adv Manuf Technol (2017) 92:2853–2865
DOI 10.1007/s00170-017-0367-1

ORIGINAL ARTICLE

Force-based tool wear estimation for milling process using Gaussian mixture hidden Markov models

Dongdong Kong[1] · **Yongjie Chen**[1] · **Ning Li**[1]

Received: 20 November 2016 / Accepted: 29 March 2017 / Published online: 11 April 2017

Abstract Tool wear monitoring system is of vital importance for the guarantee of surface integrity and manufacturing effectiveness. To overcome the weaknesses of neural networks, a new tool wear estimation model based on Gaussian mixture hidden Markov models (GMHMM) is presented. Nine types of time-domain features are extracted from the milling force signals which are obtained under four sorts of tool wear state. Besides, the sensitive features which can indicate the tool wear states accurately are selected out by correlation analysis. To test the effectiveness of the presented model, the selected sensitive features serve to identify the tool wear states by utilizing GMHMM and back-propagation neural network (BPNN), respectively. Moreover, the identification performance of GMHMM under the combinations of various numbers of Gaussian mixtures and various lengths of observation sequence is analyzed to verify the practicability of the presented tool wear model. The experimental results show that the GMHMM-based model can identify the tool wear states effectively and GMHMM outperforms the BPNN model in accuracy and stability. This method lays the foundation on tool wear monitoring in real industrial settings.

Keywords Tool wear monitoring · Milling force signals · Correlation analysis · GMHMM · BPNN

✉ Dongdong Kong
kodon007@163.com

[1] School of Mechanical Science & Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, China

1 Introduction

Tool wear degree is of vital importance for surface quality and dimensional tolerances of the workpieces during manufacturing process. Severe tool wear or tool breakage may lead to not only scrapped components but also possible damage to the machine tool. In addition, tool failure leads to at least 20% of unscheduled downtime in modern manufacturing systems [1]. However, false judgment of tool failure is of frequent occurrence since tool wear is a complex phenomenon which increases non-linearly. It is highly desirable to develop a reliable and effective monitoring system which can recognize the tool wear states real-timely so as to guarantee product quality and simultaneously reduce unexpected downtime.

There are mainly two methods for tool wear monitoring: direct and indirect methods. Direct methods [2, 3] mostly depend on machine vision which directly measures the tool wear value or area. Indirect methods [4, 5] are the most widely used techniques which perform wear estimation by establishing the corresponding models between tool wear and the related monitoring signals obtained from cutting process, such as cutting forces [6, 7], vibrations [8, 9], and acoustic emission [10–12]. The decision-making support systems, such as neural networks [13], SVM [14–17], and cluster analysis, are then utilized to recognize tool wear states. In the classification methods, neural networks (NNs) are the most widely used methods, such as artificial neural network (ANN) [18], fuzzy neural network (FNN) [19], dynamic Bayesian network (DBN) [20], multi-layer perceptron (MLP) [21], and self-organizing map (SOM) [22]. However, successful utilization of NN-based monitoring system depends heavily on proper selection of network structure and their accuracies are limited to some extent due to the connatural limitations of NNs, such as over-fitting, local minimum value, and poor generalization. D.F. Shi and N.N. Gindy [14] employed least squares support

vector machines (LS-SVM) to monitor the tool wear value during broaching process. The results showed that the predicted tool wear obtained from the predictive model had a good agreement with the experimental measurement. D.D. Kong et al. [15] utilized *v*-support vector regression (*v*-SVR) to achieve tool wear monitoring during turning process, and the results showed that the presented tool wear predictive model outperformed the back-propagation neural network (BPNN) model. Moreover, the standard SVM belongs to binary classifier which makes it especially suitable for monitoring tool breakage [16]. However, the standard SVM shows some limitations in multi-classification of tool wear states since the penalty factor and kernel parameter are hard to be determined due to the large optional range [17]. And the fine distinctions of similar features between contiguous states are hard to distinguish when the tool wear state changes from one to the next.

In this paper, a GMHMM-based monitoring system is utilized for the multi-classification of tool wear states. Hidden Markov models (HMMs) have been widely utilized in speech recognition where the signals are inherently non-stationary [23]. The application of HMMs in tool wear monitoring has made headway since the monitoring process is analogous to speech recognition in which the signal features are classified into corresponding states. Wang et al. [24] achieved the tool state detection (for sharp and worn tools) in turning process using discrete HMMs based on vibration signals and the average recognition rate reached up as high as 97%. N.N. Bhat et al. [25] utilized texture analysis and HMM technology to classify sharp, semi-dull, and dull tool in turning, with an average of 95% accuracy. C. Scheffer et al. [26] compared the performance of NNs and modified HMMs for continuous estimations of tool wear in turning. However, the estimation of a large number of parameters in the modified HMMs needs large-scale training data which make it unpractical in industry application. Cetin and Ostendorf [27] applied multi-rate HMMs to realize identification of the three states of tool wear by the vibration signal in milling process. T. Boutros and M. Liang [28] also detected correctly the state of the tool (sharp, worn, or broken) by using discrete HMM and the acoustic signal in milling. A. A. Kassim et al. [29] successfully recognized four distinct states of tool condition by the utilization of HMM based on the surface textures in milling. But the influence of cutting fluid limits the further promotion of this method in industry. H. M. Ertunc [30] and P. BARUAH [31] applied GMHMM in drilling tool monitoring and well expressed the progressive process of tool wear. However, they did not pay attention to the effect of the number of Gaussian mixtures on the performance of the model. Past research application of HMM in tool wear monitoring mainly focused on discrete HMMs or semi-continuous HMMs and achieved binary or three classification of tool wear states. In discrete HMMs, the continuous signals are converted into discrete sequences by vector quantization which will lead to loss of information [23]. Hence, previous application of HMMs does not make full use of the advantages of the method and are not suitable to the problem in this study.

The continuous HMMs has obvious advantage than the discrete HMMs in fault diagnosis since the continuous varying multi-dimensional signal features can be approximated by several probability density functions [32]. In this work, the sensitive features under four tool wear states are directly utilized to train the corresponding GMHMM model. The number of Gaussian mixtures in GMHMM determines whether the fitting of training data is sufficient. In addition, the length of observation sequence influences the identification accuracy of GMHMM for tool wear states. The identification accuracy of GMHMM under different combinations of various numbers of Gaussian mixtures and various lengths of observation sequence is analyzed to verify the practicability of the presented tool wear model and seek out the suitable combination.

The objective of this study is to identify the tool wear states in milling process by utilizing the GMHMM-based method. The paper is organized as follows. The experimental setup, data collection, and related theoretical methods are presented in Section 2. Analysis of the GMHMM method and experimental results are given in Section 3. Finally, Section 4 concludes this paper.

2 Materials and methods

2.1 Experimental setup and data collection

2.1.1 Experimental setup

This study focuses on the correlation between milling force signals and tool wear states for milling of titanium alloy (Ti-6AI-4V). The cutting tests are carried out on a CNC milling machine (Mikron UCP800Duro). The tool inserts (Walter SPMT1204AEN-WSP45) are connected to the machine spindle through a tool holder (Walter F2233.B.080.Z06.07). The experimental setup for the monitoring system is shown in Fig. 1. The signal acquisition devices consist of dynamometer (Kistler 9257A), charge amplifier (Kistler 5070A), data acquisition card (Kistler 5697A), and the DynoWare package (Kistler 2825A). The milling force signals are acquired by utilizing the dynamometer with sampling rate at 5 kHz and displayed via the DynoWare software. The cutting parameters selected from the recommended range for the inserts are listed in Table 1.

The tool flank wear is measured by using the video measuring system (VMS-1510G) which comprises a CCD camera and a 2D measurement software (QIM1008). Three tool inserts in all are symmetrically installed on the tool holder in this work. The average value $VB = 1/3(VB_1 + VB_2 + VB_3)$ of tool

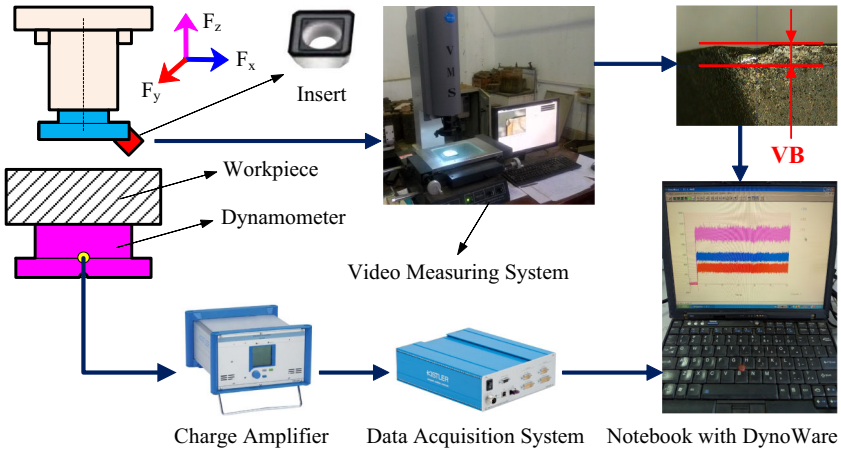


Fig. 1 The experimental setup

flank wear of the three inserts is adopted as the final tool wear value and utilized to assess tool wear state which includes four types: initial wear, normal wear, severe wear, and breakage as shown in Table 2. The inserts are detached to measure the tool flank wear by the video measuring system after each cutting process until broken. Tool wear value of the new inserts will exceed 0.1 mm when the inserts cut into workpiece due to the existence of greater impact in milling process. Besides, 0.3 mm is adopted as the threshold to determine whether the tool inserts are broken since the CNC machine will produce larger vibration once the average wear value exceeds 0.3 mm. Tool wear value of the three inserts and the corresponding tool wear states are listed in Table 3. Tool wear morphologies under four tool wear states are illustrated in Fig. 2.

2.1.2 Data collection

Signal features of the force signals change gradually with the variation of tool wear state from sharp to broken. Previous studies have shown that many time-domain features can be taken as indicators of tool wear states [33]. In this study, nine types of time-domain features are extracted from the milling force signals. The mathematical description of these features is shown in Table 4. For each cutting process, there are three mutually perpendicular milling force signals (F_x, F_y, F_z) which are collected from the dynamometer as shown in Fig. 1. There are 27 time-domain features in all could be extracted from the milling force signals. Notice that not all the extracted features can be used to identify tool wear state effectively. Besides, the insensitive features will impact the performance of classifier. Therefore, feature selection should be performed before sending the extracted features to the monitoring system so as to remove irrelevant features and improve the prediction accuracy. In this work, sensitive features that correlate well with tool wear are selected out by using correlation coefficient method expressed by Eq. (1),

$$\rho_{xy} = \frac{1}{N}\sum_{i}^{N}\left(\frac{x_i-\mu_x}{\sigma_x}\times\frac{y_i-\mu_y}{\sigma_y}\right) \tag{1}$$

where μ_x and μ_y are the mean value of x and y, respectively. σ_x and σ_y are the standard deviation of x and y, respectively. ρ_{xy} reflects the degree of linear correlation between x and y, $0 < |\rho_{xy}| < 1$. A larger absolute value $|\rho_{xy}|$ represents x has a higher correlation with y.

Firstly, feature extraction is performed from the three sets of experiment as listed in Table 1. Then, correlation analysis between the extracted features and the corresponding tool wear value is conducted in each set of experiment, respectively. The selected sensitive features (in italic) ranked in terms of average correlation coefficients $\left(|\overline{\rho}_{xy}| > 0.72\right)$ are listed in Table 5. The correlation analysis intuitively reveals the linear dependence of the signal features and tool wear degree. Note that all the signal features should be normalized to [−1, 1], so as to eliminate the mutual influence of different orders of magnitude. Spatial distribution of the selected features is illustrated in Fig. 3. It can be seen that the selected features show a

Table 1 Experimental cutting parameters

Test No.	Cutting speed (m/min)	Feed (mm/z)	Cutting depth (mm)	Cutting width (mm)
1	45	0.18	0.5	75
2	60	0.14	0.5	75
3	60	0.18	0.5	75

Table 2 Categories of tool wear state

Tool wear state	Range of tool wear value (mm)	Classification
Initial wear	0.1 ~ 0.2	1
Normal wear	0.2 ~ 0.25	2
Severe wear	0.25 ~ 0.3	3
Breakage	>0.3	4

certain degree of clustering which provides effective information for tool wear estimation. It is also explained that correlation analysis can help to select out the sensitive features.

2.2 Tool wear estimation based on GMHMM

2.2.1 The fundamentals of GMHMM

Symbols	
The standard HMMs mode	
$O=\{o_1, o_2, \cdots, o_T\}$	the observation sequence
$I=\{i_1, i_2, \cdots, i_T\}$	the state sequence
N	the number of hidden states in the model
M	the number of observable symbols in each state
$\pi=\{\pi_i\}$	initial state probability vector
$A=\{a_{ij}\}$	state transition probability matrix
$B=\{b_j(k)\}$	observation probability matrix
$\lambda=(\pi, A, B)$	the complete description for an HMM

Hidden Markov models (HMMs) are an extension of Markov chains [34]. The HMMs is a double-layered stochastic process: one is the transition from one state to another state which is invisible, the other is the output symbol generated at each state which is observable. There are three basic problems for HMMs:

1. Probability calculation for $P(O|\lambda)$, given the model $\lambda=(\pi, A, B)$ and observation sequence $O=\{o_1, o_2 \cdots o_T\}$. The forward-backward algorithm is effective for this problem.
2. Model estimation for $\lambda=(\pi, A, B)$ so as to maximize $P(O|\lambda)$, given the observation sequence $O=\{o_1, o_2 \cdots o_T\}$. The Baum-Welch algorithm is the problem-solving method.
3. Maximization of $P(I|O)$, i.e., state sequence estimation, given the model $\lambda=(\pi, A, B)$ and observation sequence $O=\{o_1, o_2 \cdots o_T\}$. The Viterbi algorithm is the common solution.

Table 3 Tool wear value of the three inserts and the corresponding tool wear state (Test No. 1)

VB (mm)	Initial wear	Normal wear	Severe wear	Breakage
Insert 1	0.1330	0.2300	0.2850	0.3420
Insert 2	0.1460	0.2470	0.3040	0.3500
Insert 3	0.1410	0.2400	0.3030	0.3630
Average value	0.1400	0.2390	0.2973	0.3517

HMMs can be classified as discrete HMMs and continuous HMMs, depending on whether the observation sequences are discrete or continuous. In the continuous HMMs, the continuous probability density function for state j can be described as a weighted sum of K Gaussian mixtures [23],

$$b_j(O)=\sum_{k=1}^{K} c_{jk} b_{jk}(O)=\sum_{k=1}^{K} c_{jk} N\left(O, \mu_{jk}, U_{jk}\right) \tag{2}$$

where O is the observation sequence being modeled, K is the number of Gaussian mixtures for state j, c_{jk} is the weight coefficient for the kth mixture in state j, and $N(O, \mu_{jk}, U_{jk})$ is a Gaussian density function with mean vector μ_{jk} and covariance matrix U_{jk} for the kth mixture in state j. The weight coefficient c_{jk} must satisfy the stochastic constraints, i.e.,

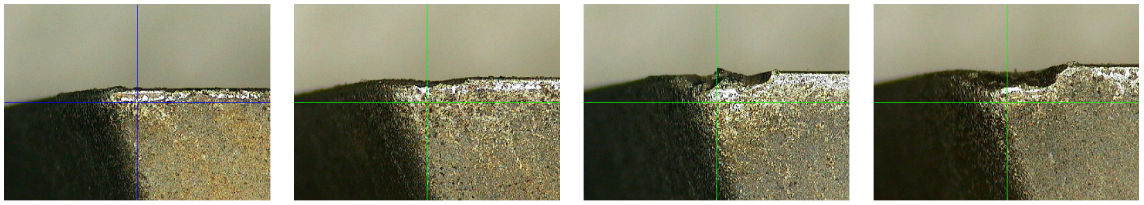
$$c_{jk} \geq 0, \quad 1 \leq j \leq N, 1 \leq k \leq K \tag{3}$$

$$\sum_{k=1}^{K} c_{jk}=1, \quad 1 \leq j \leq N \tag{4}$$

In this way, the Gaussian mixture hidden Markov models (GMHMM) with a continuous output can be expressed by $\lambda=(\pi, A, c_{jk}, \mu_{jk}, U_{jk})$. In this work, the model parameters λ in GMHMM need to be estimated and trained, i.e., adjusting the GMHMM parameters λ so as to maximize the probability of the observation sequence $P(O|\lambda)$. The model parameters λ can be estimated iteratively by the well-known Baum-Welch algorithm base on the training data. Multiple observation sequences are utilized for the training of GMHMM in this study. Assuming that the training data contain S observation sequences of length T, which is defined as $\{O_1, O_2, \cdots, O_S\}$. Firstly, an initial guess of a set of appropriate parameters λ for GMHMM need to be carried out. The new parameters $\overline{\lambda}$ for GMHMM are then calculated by iteration procedures as following [35]:

$$\overline{\pi}_i=\sum_{s=1}^{S}\sum_{j=1}^{N} \varepsilon_1^s(i,j) \tag{5}$$

$$\overline{a}_{ij}=\frac{\sum_{s=1}^{S}\sum_{t=1}^{T-1}\varepsilon_t^s(i,j)}{\sum_{s=1}^{S}\sum_{t=1}^{T-1}\sum_{n=1}^{N}\varepsilon_t^s(i,n)} \tag{6}$$



Initial wear (0.133mm) Normal wear (0.23mm) Severe wear (0.285mm) Breakage (0.342mm)

Fig. 2 Tool wear morphologies under different tool wear states (Test No. 1: insert 1)

$$\overline{c}_{jk} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\gamma_t^s(j,k)}{\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{k=1}^{K}\gamma_t^s(j,k)} \tag{7}$$

$$\overline{\mu}_{jk} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\gamma_t^s(j,k)\cdot o_t^s}{\sum_{s=1}^{S}\sum_{t=1}^{T}\gamma_t^s(j,k)} \tag{8}$$

$$\overline{U}_{jk} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\gamma_t^s(j,k)\cdot\left(o_t^s-\mu_{jk}\right)\left(o_t^s-\mu_{jk}\right)'}{\sum_{s=1}^{S}\sum_{t=1}^{T}\gamma_t^s(j,k)} \tag{9}$$

where $\varepsilon_t(i,j)$ represents the probability of state i transfers to state j at time t. $\gamma_t(j,k)$ is the probability output of the kth mixture component being in state j at time t, given the model parameters λ and the observation sequence O, i.e.,

$$\gamma_t(j,k) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\cdot\frac{c_{jk}N\left(o_t,\mu_{jk},U_{jk}\right)}{\sum_{k=1}^{K}c_{jk}N\left(o_t,\mu_{jk},U_{jk}\right)} \tag{10}$$

where $\alpha_t(j)$ and $\beta_t(j)$ are the forward and backward probability, respectively.

The iteration procedures should be carried out continuously until the increment meets the convergence condition which is the final estimated model $\overline{\lambda} = \left(\overline{\pi},\overline{A},\overline{c}_{jk},\overline{\mu}_{jk},\overline{U}_{jk}\right)$ should satisfy the inequality that $|P(O|\lambda^{i+1}) - P(O|\lambda^{i})| < \varepsilon$. Note that initial guess of the model parameters λ is important for obtaining a suitable GMHMM model since the Baum-Welch algorithm only guarantees local optimal value.

Table 4 Mathematical description of the time-domain features

Signal features	Mathematical expression
Mean $(\overline{F_x}, \overline{F_y}, \overline{F_z})$	$\mu = E(\lvert x_i\rvert)$
Maximum value (Max)	$x_{\mathrm{Max}} = max(\lvert x_i\rvert)$
Peak to valley (PV)	$x_{PV} = x_{\mathrm{Max}} - x_{Min}$
Root mean square (Rms)	$x_{Rms} = \{E(x_i^2)\}^{1/2}$
Standard deviation (Std)	$x_{Std} = \sigma = \{E[(\lvert x_i\rvert - \mu)^2]\}^{1/2}$
Skewness (Ske)	$x_{Ske} = E\{[(\lvert x_i\rvert - \mu)/\sigma]^3\}$
Kurtosis (Kur)	$x_{Kur} = E\{[(\lvert x_i\rvert - \mu)/\sigma]^4\}$
Form factor (Fmf)	$x_{Fmf} = x_{Rms}/\mu$
Force ratio	$\overline{F_x}/\overline{F_z}$, $\overline{F_x}/\overline{F_y}$, $\overline{F_y}/\overline{F_z}$

2.2.2 Multi-classifier based on GMHMM

Supervised learning is made up of two aspects, i.e., learning of the labeled data and unlabeled data classification. In this work, there are four tool wear states to be monitored and each state is modeled by a separate GMHMM model. The training data and test data for each tool wear state are reorganized as the observation sequences. The GMHMM-based monitoring process contains the following two steps: Firstly, we need to build a GMHMM model λ_i $(i = 1, 2, 3, 4)$ for each tool wear state as shown in Table 2. The model parameters $\lambda_i = (\pi, A, c_i, \mu_i, U_i)$ need to be estimated so as to maximize the probability output $P(O|\lambda_i)$ of the observation sequences that obtained from the training data for the ith tool wear state. The Baum-Welch algorithm is utilized to estimate the model parameters of GMHMM as shown in Section 2.2.1. Then, the trained model $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ serves to identify the corresponding tool wear states of the observation sequences in test data. The observation sequence of test data is sent to the four trained models $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ at the same time and the log-likelihood probability output for each tool wear model is computed. The model with the maximum output is selected to be the corresponding tool state as expressed by Eq. (11). The flowchart of GMHMM for tool wear state identification is illustrated in Fig. 4.

$$i' = \arg\max_{1\le i\le 4} \log P(O|\lambda_i) \tag{11}$$

3 Experimental results and analysis

This paper aims at constructing the theoretical models between tool wear states and feature signals that facilitates the control of machining quality and makes full use of tool life. The tool wear state is evaluated quantitatively by the mean value of tool flank wear as shown in Tables 2 and 3. To reflect the tool wear process as accurately as possible, nine types of time-domain features under four sorts of tool wear state are extracted from the milling force signals as listed in Table 4. Besides, the correlation analysis is employed to select out the sensitive features that correlate well with tool wear degree. The selected time-domain features change gradually with the worsening of tool wear degree. The 18 sensitive features as

Table 5 The selected sensitive features

Signal features	$\overline{F_x}$	$\overline{F_y}$	$\overline{F_z}$	Max_F_x	Max_F_y	Max_F_z	PV_F_x	PV_F_y	PV_F_z
$\overline{\rho}_{xy}$	*0.8620*	*0.8975*	*0.8805*	*0.7982*	*0.8402*	*0.8515*	*0.7983*	*0.8369*	*0.7898*
Signal features	Rms_F_x	Rms_F_y	Rms_F_z	Std_F_x	Std_F_y	Std_F_z	Ske_F_x	Ske_F_y	Ske_F_z
$\overline{\rho}_{xy}$	*0.8675*	*0.8971*	*0.8797*	*0.8597*	*0.8365*	*0.8244*	0.6220	−0.0648	−0.0262
Signal features	Kur_F_x	Kur_F_y	Kur_F_z	Fmf_F_x	Fmf_F_y	Fmf_F_z	$\overline{F_x}/\overline{F_z}$	$\overline{F_y}/\overline{F_z}$	$\overline{F_x}/\overline{F_y}$
$\overline{\rho}_{xy}$	0.1002	0.4226	−0.2649	0.7157	*−0.7765*	−0.5550	*−0.8259*	−0.1211	*−0.8677*

listed in Table 5 and the corresponding 4 cutting parameters make up the 22-dimension feature vectors which are reorganized to be the observation sequences. Note that the signal features in feature vectors need to be normalized to [−1, 1] before sent to GMHMM in case that large order of magnitude weakens the effect of small one. To analyze and verify the reliability of the GMHMM-based model, 200 sample feature vectors are extracted from each tool wear state, respectively. The extracted sample feature vectors are divided into training data and test data which have the same size and

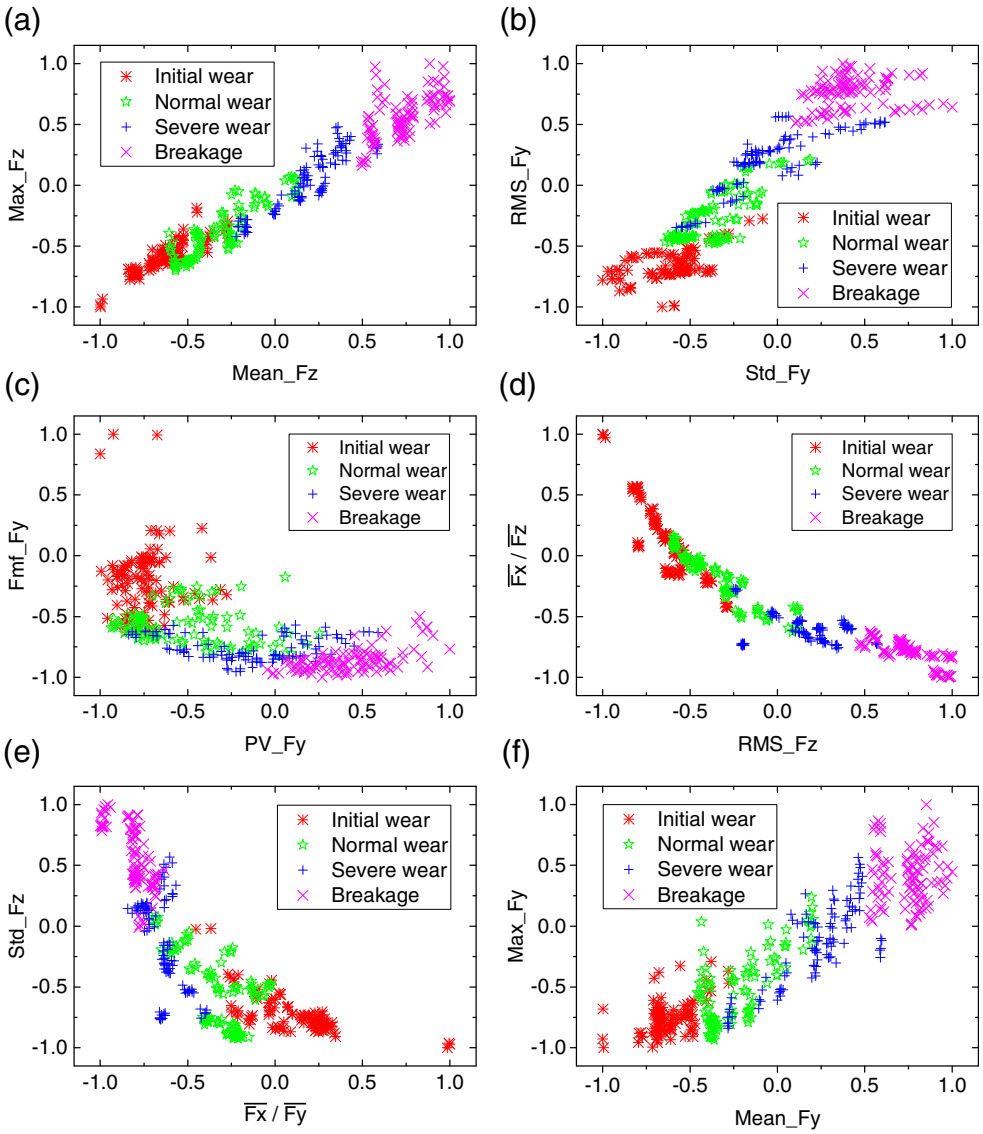


Fig. 3 Samples of spatial distribution for the selected features

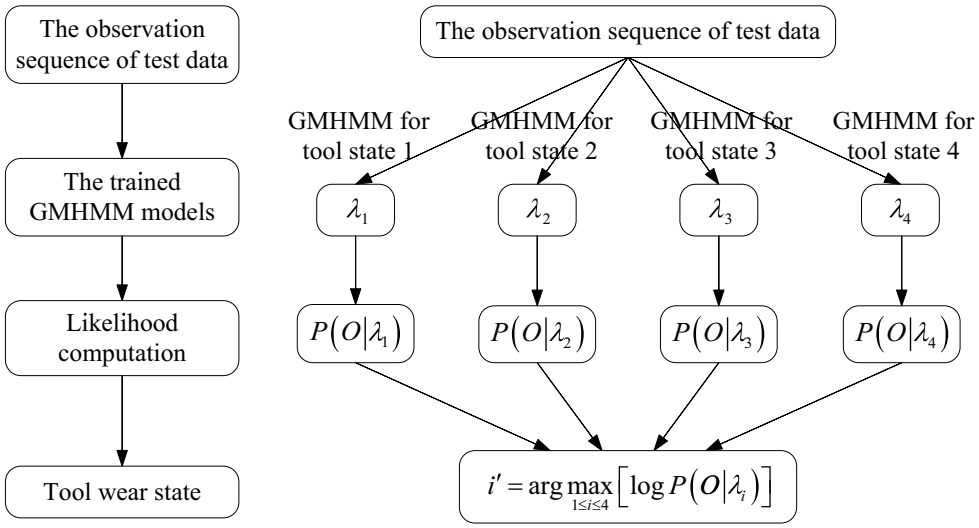


Fig. 4 The flowchart of GMHMM for tool wear estimation

do not contain each other. The observation sequences of the training data under four tool wear states are utilized to train the GMHMM-based model $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, respectively. As soon as the training stage is finished, the tool wear state estimation model is produced and can be utilized to identify the tool wear state of the observation sequence obtained from the test data.

3.1 Determination of the GMHMM structure

There are diverse topology structures for HMMs [23]. The left-right HMM is selected in this work since tool wear is always increasing during the cutting process and cannot go back. The tool wear starts from initial wear state, goes through normal wear state and severe wear state, finally ends up in tool

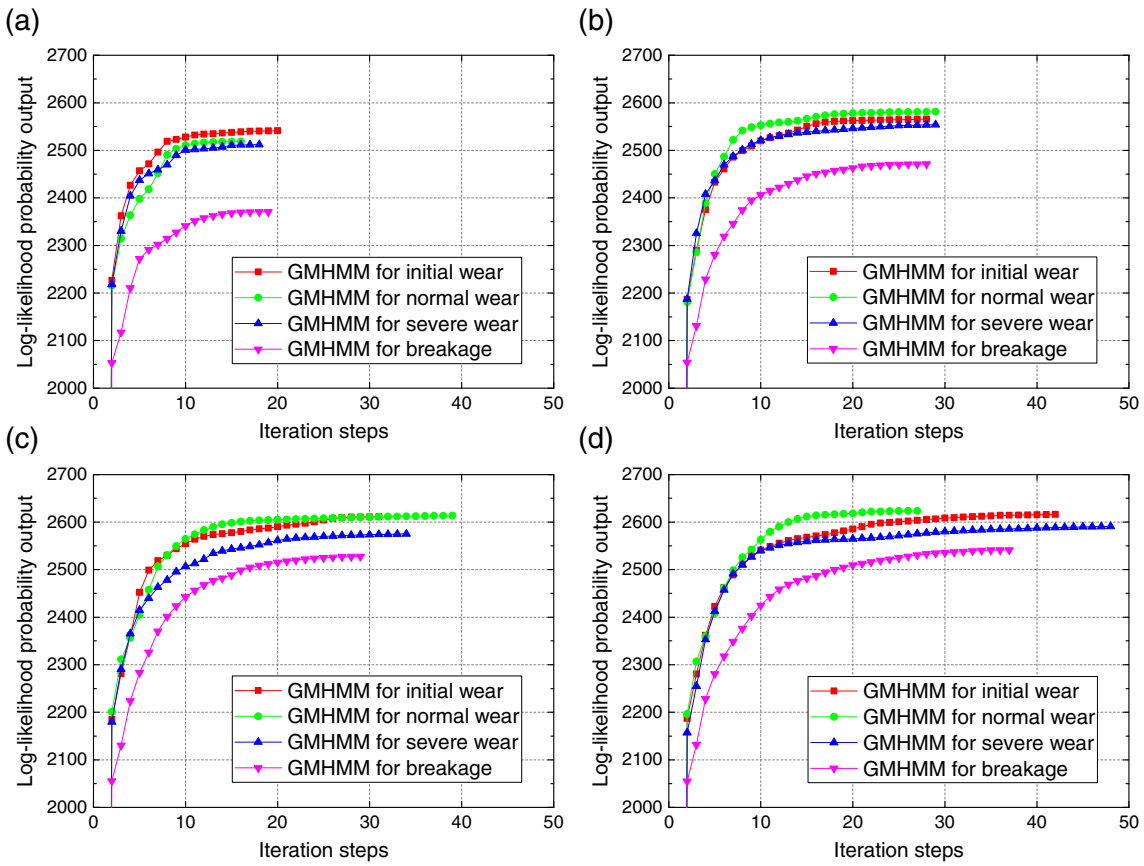


Fig. 5 The training process of GMHMM under four sorts of tool wear state. (**a**) $K = 2,\ T = 3$; (**b**) $K = 4,\ T = 3$; (**c**) $K = 6,\ T = 3$; (**d**) $K = 8, T = 3$

Table 6 The consuming time for the training of the GMHMM-based model (s)

	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$
$T=2$	6.900	12.209	18.673	24.460	26.080	35.298	35.769
$T=3$	5.137	8.263	10.798	15.886	20.912	24.471	30.133
$T=4$	3.798	6.608	8.855	13.932	14.848	17.461	20.875
$T=5$	3.200	6.789	9.903	9.823	14.180	14.363	17.971
$T=6$	2.691	3.996	7.314	9.571	10.332	13.600	16.039
$T=7$	2.681	3.654	6.490	7.452	10.252	11.934	15.567
$T=8$	2.402	4.666	4.614	6.723	7.476	9.935	10.426

Note: K represents the number of Gaussian mixtures; T represents the length of observation sequence

breakage. Therefore, the left-right HMM is especially suitable for TCM since it can only transfer from the former state to the later state and there is no leapfrog jump between states. Once the topology of HMMs is selected, the number of hidden states in HMMs also needs to be determined. Note that the hidden states in HMMs and the tool wear states do not correspond one-to-one. The hidden states in HMMs are only the transition states within HMMs and lack clear physical meaning [36]. There are no fixed criteria for the determination of the number of hidden states in HMMs.

In this study, the visible contents of the tool wear state transitions are bestowed upon the GMHMM-based model for ease of understanding and reducing the complexity of the model. Although π and A have an important influence on the Markov chain, their initial value has little effect on the final clustering results [37]. In this way, π and A can be randomly or uniformly initialized. The uniform value is adopted for the initialization of π and A in this work so as to guarantee the stability of GMHMM. The initial state probability of the GMHMM-based model is $\boldsymbol{\pi} = [0.25 \;\; 0.25 \;\; 0.25 \;\; 0.25]$. As for the state transition probability matrix $\boldsymbol{A} = \{a_{ij}\}$ of the GMHMM-based model, a_{ij} are initialized as follows since the insert will go through four sorts of tool wear state.

$$A = \{a_{ij}\} = \begin{bmatrix} \text{state1} & \text{state2} & \text{state3} & \text{state4} \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Another one is the number of Gaussian mixtures K. Mixtures of Gaussians are utilized to model the output of signal features in this study. The number of Gaussian mixtures directly affects the accuracy of data fitting and the identification rate for unknown data. The number of matrices (c_i, μ_i, U_i) contained in the Gaussian mixture model is $(N+1)\cdot K+1$. The computation load in training process for the GMHMM-based model will increase gradually with the increment of the number of Gaussian mixtures K. Less number of Gaussian mixtures will lead to inadequate description of the training data. Increasing the number of Gaussian mixtures is helpful to improve the identification rate in the case of sufficient training samples. However, the computation load will also increase

Table 7 Samples of log-likelihood output of the observation sequences in test data (Test No. 1)

Case No.	Initial wear model $\overline{\lambda}_1$	Normal wear model $\overline{\lambda}_2$	Severe wear model $\overline{\lambda}_3$	Tool breakage model $\overline{\lambda}_4$	Actual state	Predicted state
1	*78.133*	−14.494	−192.586	−561.031	1	1
2	*78.894*	−12.519	−188.158	−561.612	1	1
3	*77.028*	−3.429	−167.325	−513.630	1	1
4	43.220	*71.070*	−45.089	−322.500	2	2
5	40.370	*71.789*	−32.277	−342.084	2	2
6	35.434	*75.385*	19.418	−366.385	2	2
7	−61.108	61.907	*73.459*	−120.517	3	3
8	−60.973	58.840	*78.333*	−105.909	3	3
9	−96.551	54.377	*79.005*	−78.836	3	3
10	−493.881	−29.314	44.230	*70.587*	4	4
11	−619.266	−46.604	22.930	*70.490*	4	4
12	−763.058	−108.766	0.374	*72.763*	4	4

Note: the GMHMM-based models $\{\overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3, \overline{\lambda}_4\}$ are obtained under $K=3$ and $T=3$

accordingly and the training process may not converge. Thus, the selection of the number of Gaussian mixtures is of vital importance for the GMHMM-based model.

Moreover, the length of observation sequence T also influences the identification rate of the GMHMM-based model. The identification rate will inevitably increase with the

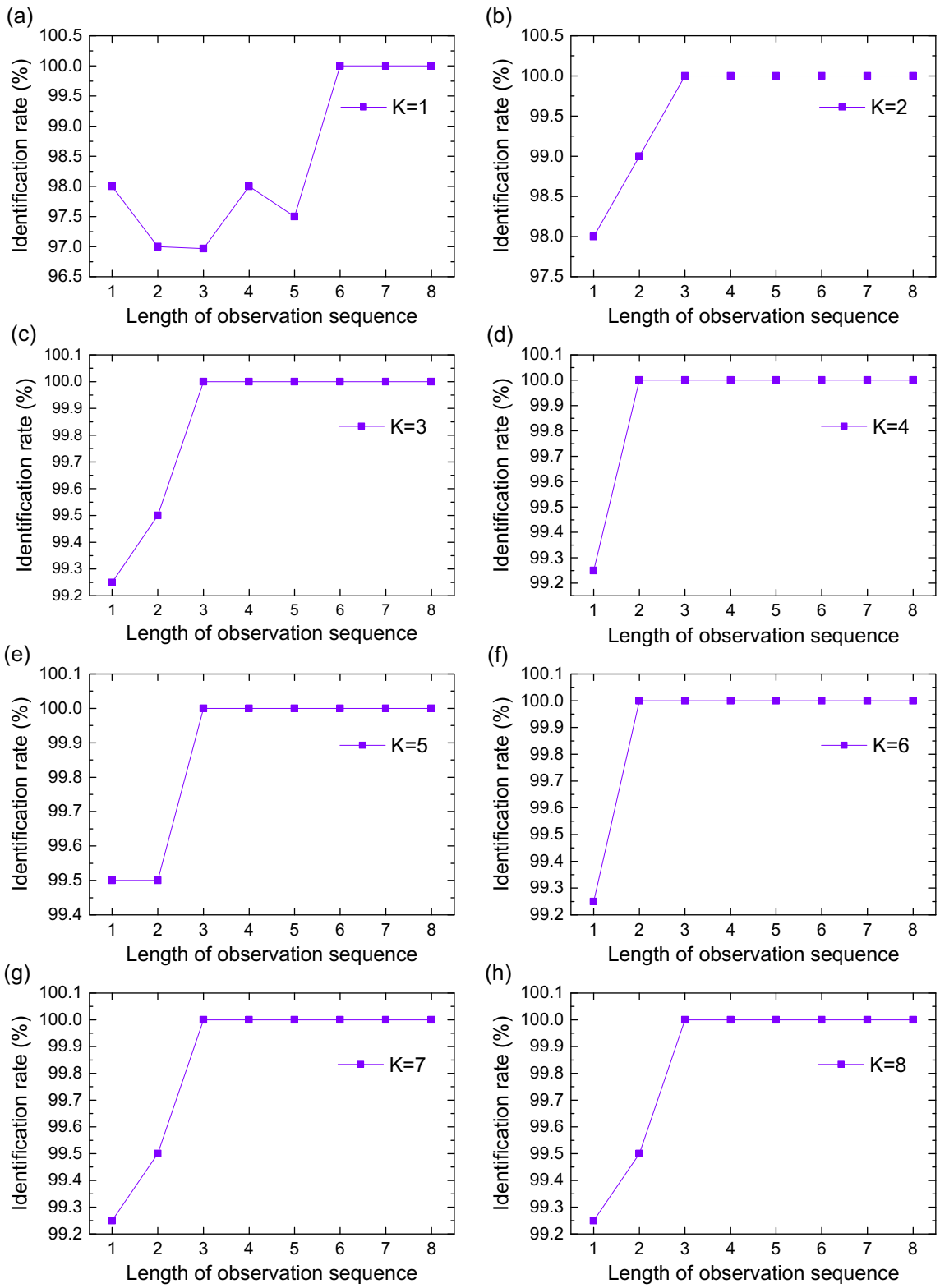


Fig. 6 The identification rate of the GMHMM-based monitoring system

increment in the length of observation sequence [38]. However, the shorter the length of observation sequence, the less will be the time-cost of tool wear estimation. The monitoring system can detect tool breakage as soon as possible so as to ensure machining quality and protect the manufacturing system. The classification performance of the GMHMM-based model under different combinations of various numbers of Gaussian mixtures and various lengths of observation sequence need to be analyzed so as to select out the appropriate combination of K and T.

3.2 GMHMM for tool wear estimation

The model parameters $\lambda_i = (\pi, A, c_i, \mu_i, U_i)$ need to be estimated as soon as the structure of GMHMM is determined. The training and testing process are carried out by using the Hidden Markov model (HMM) Toolbox for Matlab [39]. There are four GMHMM-based models that need to be trained as shown in Fig. 4 which corresponds to four sorts of tool wear state, respectively. The training for the GMHMM-based models $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ is carried out by the Baum-Welch algorithm as presented in Section 2.2.1. Parameter initialization is important for obtaining a suitable GMHMM model since the Baum-Welch algorithm only guarantees local optimal value. The initialization of π and A are presented in Section 3.1. In addition, the K-means method [40] is utilized for the initial guess of the model parameters (c_i, μ_i, U_i) before the training. Maximum number of iterations during the training process is set to 100. And the convergence threshold $\varepsilon = 0.0001$ is set for the end of training. Each GMHMM-based model λ_i is trained with its corresponding observation sequences obtained from the training data. The training process of GMHMM under four sorts of tool wear state are illustrated in Fig. 5. It is obvious that when the length of observation sequence T is fixed, the iteration steps show an increasing trend with the increment of the number of Gaussian mixtures K. Besides, time-cost of the training process also increases accordingly due to the raise of single iteration time. The consuming time for the training of the GMHMM-based model under different combinations of K and T is listed in Table 6.

Once the training process is finished, the multi-classifier based on GMHMM is produced and can be utilized to identify the tool wear state of the observation sequence obtained from the test data. The observation sequences in test data are input to the four trained models $\{\overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3, \overline{\lambda}_4\}$ and the log-likelihood for each model is computed. The model with the maximum output is selected to be the corresponding tool state as shown in Fig. 4. The log-likelihood output of the observation sequences in test data is listed in Table 7. The log-likelihood output in italic represents the most probable tool state. The identification rate of the GMHMM-based monitoring system for four sorts of tool wear state is shown in Fig. 6. When there is only one Gaussian function to model the output, GMHMM degenerate into Gaussian-HMM whose performance is inferior to GMHMM as shown in Fig. 6(a). The experimental results show that increment in the length of observation sequence can guarantee the stability of the monitoring system. However, longer observation sequence needs longer period for feature extraction which is detrimental to real-time monitoring of tool wear. It can be seen that the sequence of three observations ($T = 3$) is enough to estimate the tool wear state effectively in machining process. Besides, the GMHMM with three Gaussian mixtures ($K = 3$) is enough for the GMHMM-based monitoring system considering the fitting effect of training data and the training time. Consequently, the combination of $K = 3$ and $T = 3$ is a better choice for the GMHMM-based monitoring system so as to guarantee the system performance and stability without sacrificing the training time. A complete tool wear monitoring process for test no. 1 is illustrated in Fig. 7. It can be seen that

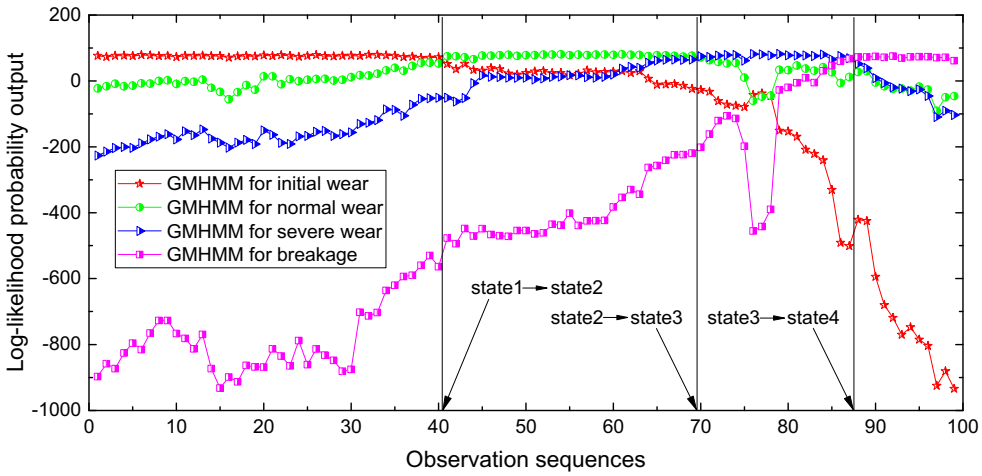


Fig. 7 A complete tool wear monitoring process for Test No. 1 ($K = 3, T = 3$)

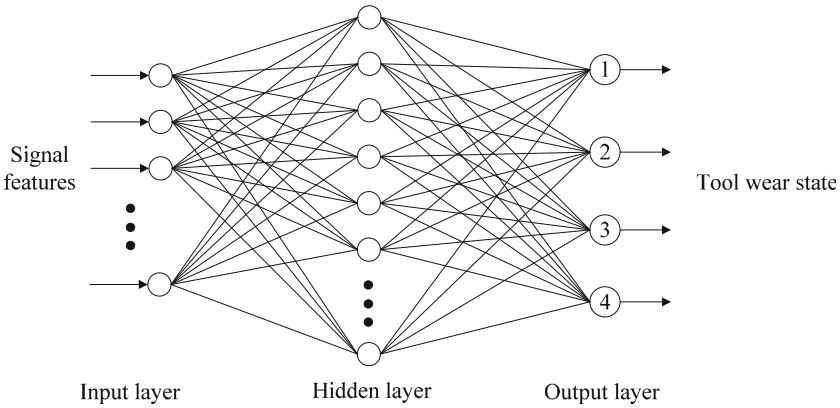


Fig. 8 Structure of the BPNN model for tool wear estimation

when the tool is in a certain wear state, the log-likelihood output of the corresponding wear model is maximum. The maximum log-likelihood output fluctuates narrowly in the four models $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, which reveals the stability of the GMHMM-based monitoring system. The result indicates that the GMHMM-based tool wear estimation model can identify the tool wear state effectively in machining process.

The GMHMM-based monitoring system can also be applied to various combinations of tool inserts and materials since the presented method is subject to the sensitive features extracted from the milling process. Similarly, other machining processes such as turning and drilling can also be monitored by utilizing the GMHMM-based method, provided that the corresponding sensitive features can be found.

3.3 Comparison with BPNN model

To demonstrate the advantages of the GMHMM model as discussed above, back-propagation neural network (BPNN) is adopted to identify tool wear states. Structure of the BPNN model for tool wear estimation is illustrated in Fig. 8. The network structure is specified as below. The linear and sigmoid activation functions are designated for the hidden layer and output layer, respectively. Twenty-five nodes are designated in the hidden layer which is definitely determined

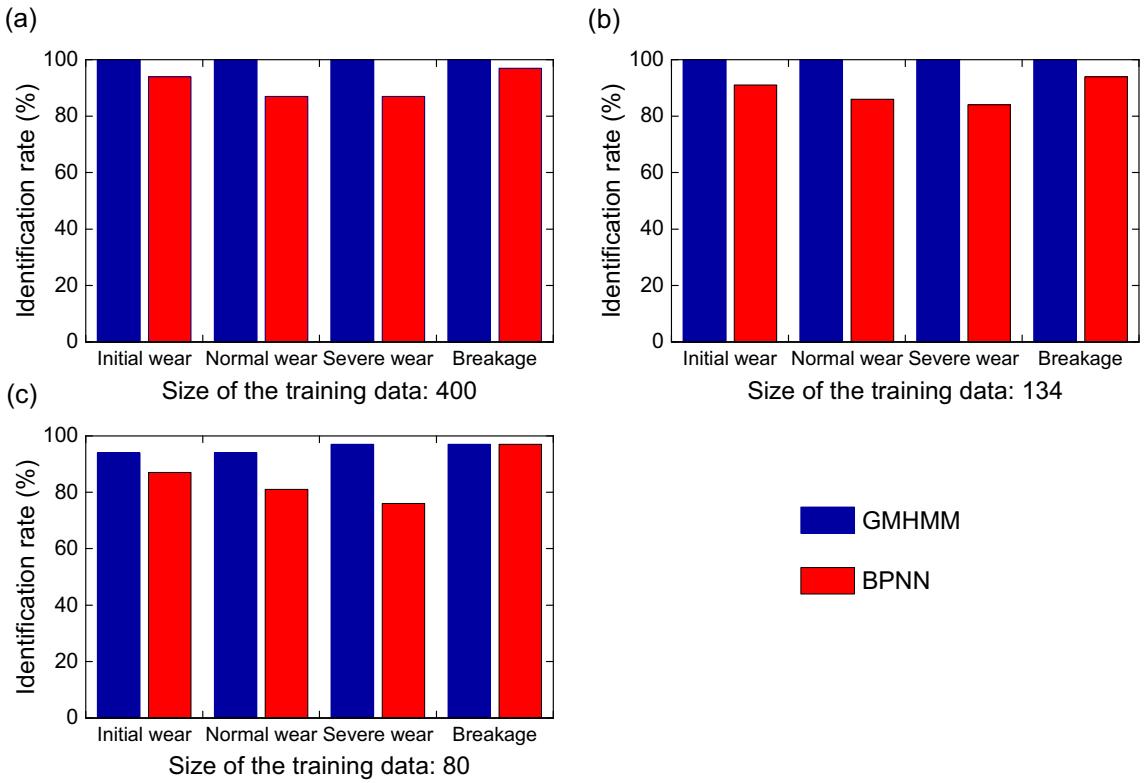


Fig. 9 Performance comparison of GMHMM and BPNN under different size of training data ($K = 3$, $T = 3$)

by trial and error. The gradient descent momentum is designated as the training algorithm in which the learning rate is set to 0.05. The maximum number of iterations is set to 10,000. The same sample data used in GMHMM are utilized in the BPNN model. Training data is utilized to train the BPNN model. Once the training process is finished, the trained BPNN model can be utilized to identify the tool wear state of test samples. Performance comparison of GMHMM and BPNN under different size of training data is presented in Fig. 9. It can be seen that the GMHMM-based model still maintain a higher identification rate even if the size of training data is sharply reduced. Besides, the variation range of GMHMM is less than BPNN with the rapid reduction of training data, which shows that the multi-classifier based on GMHMM has stronger stability and robustness. Moreover, the network structure of BPNN is difficult to determine and the algorithm converges slowly. The determination of model parameters in BPNN needs to be carried out by trial and error, which will take a considerable amount of time in comparison with GMHMM. It can be concluded that the GMHMM-based monitoring system shows a significant superiority over BPNN for tool wear estimation.

4 Conclusions

This paper presents a new tool wear estimation model based on Gaussian mixture hidden Markov models (GMHMM). To verify the practicability of the GMHMM-based model, 18 time-domain features that extracted from the milling force signals are selected as the sensitive features to identify the tool wear states. Both the number of Gaussian mixtures K and the length of observation sequences T directly affect the performance of the GMHMM-based model. The experimental results show that the combination of $K = 3$ and $T = 3$ is a better candidate for the GMHMM-based model since both the system performance and stability are guaranteed without sacrificing the training time. The identification rate of the GMHMM-based monitoring system for four sorts of tool wear state can reach up to 100%. Comparison results show that GMHMM outperforms the BPNN model in performance and stability. Besides, the selection of network structure and determination of model parameters for BPNN need to be carried out by trial and error, which will take a considerable amount of time in comparison with GMHMM. Consequently, the GMHMM-based monitoring system shows great superiority over BPNN for tool wear estimation. This method lays the foundation on tool wear monitoring in real industrial settings.

Acknowledgements The authors are grateful to the financial sponsorship from 863 National High-Tech Research and Development Program of China (No.2013AA041108).

References

1. Kurada S, Bradley C (1997) A review of machine vision sensors for tool condition monitoring. Comput Ind 34:55–72
2. Jurkovic J, Korosec M, Kopac J (2005) New approach in tool wear measuring technique using CCD vision system. International Journal of Machine Tools & Manufacture 45:1023–1030
3. Castejón M, Alegre E, Barreiro J, Hernández LK (2007) On-line tool wear monitoring using geometric descriptors from digital images. International Journal of Machine Tools & Manufacture 47: 1847–1853
4. Dimla DE Snr (2000) Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. International Journal of Machine Tools & Manufacture 40:1073–1098
5. Salgado DR, Alonso FJ (2007) An approach based on current and sound signals for in-process tool wear monitoring. International Journal of Machine Tools & Manufacture 47:2140–2152
6. Kious M, Ouahabi A, Boudraa M, Serra R, Cheknane A (2010) Detection process approach of tool wear in high speed milling. Measurement 43:1439–1446
7. Dimla DE Snr, Lister PM (2000) On-line metal cutting tool condition monitoring. I: force and vibration analyses. International Journal of Machine Tools & Manufacture 40:739–768
8. Alonso FJ, Salgado DR (2008) Analysis of the structure of vibration signals for tool wear detection. Mech Syst Signal Process 22: 735–748
9. Kilundu B, Dehombreux P, Chiementin X (2011) Tool wear monitoring by machine learning techniques and singular spectrum analysis. Mech Syst Signal Process 25:400–415
10. Li XL (2002) Acoustic emission methods for tool wear monitoring during turning. International Journal of Machine Tools & Manufacture 42:157–165
11. Liang SY, Dornfeld DA (1989) Tool wear detection using time series analysis of acoustic emission. Journal of Engineering for Industry 111(3):199–205
12. Tansel I, Trujillo M, Nedbouyan A et al (1998) Micro-end-milling—III. Wear estimation and tool breakage detection using acoustic emission signals. International Journal of Machine Tools & Manufacture 38:1449–1466
13. Sick B (2002) On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. Mech Syst Signal Process 16(4):487–546
14. Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. Mech Syst Signal Process 21: 1799–1814
15. Kong DD, Chen YJ, Li N, Tan SL (2017) Tool wear monitoring based on kernel principal component analysis and v-support vector regression. Int J Adv Manuf Technol 89:175–190
16. Li N, Chen YJ, Kong DD, Tan SL (2016) Force-based tool condition monitoring for turning process using v-support vector regression. Int J Adv Manuf Technol. doi:10.1007/s00170-016-9735-5
17. Wang GF, Yang YW, Zhang YC, Xie QL (2014) Vibration sensor based tool condition monitoring using support vector machine and locality preserving projection. Sensors Actuators A 209:24–32
18. Kaya B, Oysu C, Ertunc HM (2011) Force-torque based on-line tool wear estimation system for CNC milling of Inconel 718 using neural networks. Adv Eng Softw 42:76–84
19. Chungchoo C, Saini D (2002) On-line tool wear estimation in CNC turning operations using fuzzy neural network model. International Journal of Machine Tools & Manufacture 42:29–40
20. Tobon-Mejia DA, Medjaher K, Zerhouni N (2012) CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. Mech Syst Signal Process 28:167–182
21. Dimla DE Sr, Lister PM (2000) On-line metal cutting tool condition monitoring. II: tool-state classification using multi-layer perceptron

neural networks. International Journal of Machine Tools & Manufacture 40:769–781
22. Yen CL, Lu MC, Chen JL (2013) Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting. Mech Syst Signal Process 34:353–366
23. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
24. Wang LT, Mehrabi MG, Elijah KA (2002) Hidden Markov model-based tool wear monitoring in turning. J Manuf Sci Eng 124(3): 651–658
25. Bhat NN, Dutta S, Pal SK, Pal S (2016) Tool condition classification in turning process using hidden Markov model based on texture analysis of machined surface images. Measurement 90:500–509
26. Scheffer C, Engelbrecht H, Heyns PS (2005) A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear. Neural Comput & Applic 14:325–336
27. Cetin O, Ostendorf M (2004) Multi-rate hidden Markov models and their application to machine tool-wear classification. The IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2004 5(6):837–840
28. Boutros T, Liang M (2011) Detection and diagnosis of bearing and cutting tool faults using hidden Markov models. Mech Syst Signal Process 25:2102–2124
29. Kassim AA, Zhu M, Mannan MA (2006) Tool condition classification using hidden Markov model based on fractal analysis of machined surface textures. Mach Vis Appl 17:327–336
30. Errtunc HM, Looparo KA, Ocak H (2001) Tool wear condition monitoring in drilling operations using hidden Markov modes (HMMs). International Journal of Machine Tools & Manufacture 41:1363–1384
31. Baruah P, Chinnam RB (2005) HMMs for diagnostics and prognostics in machining processes. Int J Prod Res 43(6):1275–1293
32. Bunks C, Mccarthy D, Al-Ani T (2000) Condition-based maintenance of machines using hidden Markov models. Mech Syst Signal Process 14(4):597–612
33. Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. CIRP Ann Manuf Technol 59: 717–739
34. Wang F, Tan S, Shi HB (2015) Hidden Markov model-based approach for multimode process monitoring. Chemom Intell Lab Syst 148:51–59
35. Bilmes JA (1998) A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. International Computer Science Institute TR-97-021
36. Zhu KP, Wong YS, Hong GS (2009) Multi-category micro-milling tool wear monitoring with continuous hidden Markov models. Mech Syst Signal Process 23:547–560
37. Han JQ, Zhang L, Zheng TR (2002) Speech signal processing (2nd edition) [M]. Tsinghua University press, Beijing (in Chinese)
38. Rabiner LR, Juang BH (1993) Fundamentals of speech recognition (1st edition) [M]. Prentice Hall, New Jersey
39. Hidden Markov Model (HMM) Toolbox written by Kevin Murphy (1998). <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>
40. Ge Y, Chen QG, Jiang M, Huang YQ (2014) SCHMM-based modeling and prediction of random delays in networked control systems. J Frankl Inst 351:2430–2453