Int J Adv Manuf Technol (2017) 92:3647–3657
DOI 10.1007/s00170-017-0404-0

ORIGINAL ARTICLE

Hidden semi-Markov model-based method for tool wear estimation in milling process

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

Received: 2 November 2016 / Accepted: 3 April 2017 / Published online: 2 May 2017

Abstract This paper presents a new method for tool wear estimation in milling process by utilizing the hidden semi-Markov model (HSMM). HSMM differs greatly from the standard hidden Markov model (HMM) in state duration distribution. The model structure and corresponding parameters of HSMM can be easily determined without optimization. Two groups of experiments are carried out to prove the effectiveness of the HSMM-based method by recurring to the Gamma distribution. Five types of time-domain features that characterize tool wear states are extracted from the cutting force signals during milling process. The extracted signal features are utilized to realize tool wear estimation by means of HSMM and some other published methods, respectively. The experimental and analytical results show that the HSMM-based method can reach higher accuracy for tool wear estimation. Besides, the consuming time of HSMM for the identification of tool wear state is less than 0.05 s, which makes tool wear monitoring in industrial environment become more realistic and operable.

Keywords Tool wear estimation · Cutting force signals · Hidden semi-Markov model

1 Introduction

High productivity and high machining accuracy have become important benchmarking of the manufacturing level of a country in intelligent manufacturing era. Severe tool wear and breakage will lead to interruption of machining process, useless products, and damage to the machine-tools. Twenty percent or so machine downtime is caused by excessive tool wear in the actual industrial processes [1]. Note that the latest real-time monitoring technology for tool wear can effectively improve the utilization rate of machine tools up to 50%, increase productivity by 35%, and reduce production costs by 30% [2, 3]. Therefore, accurate estimation of tool wear states is extremely significant for the guarantee of stable manufacturing process and excellent product quality.

The complex non-linear phenomenon of tool wear makes the precise monitoring of tool wear states rather difficult. Neural networks [4, 5] and SVM [6, 7] have been the most widely utilized methods to realize tool wear estimation. Besides, hidden Markov model (HMM) is also applied to tool condition monitoring (TCM) for various machining fields such as turning [8–10], milling [11–13], and drilling [14, 15]. Rabiner [16] gives a detailed review on the theory and evolution process of the HMMs. Although HMM has been applied to TCM, there are still some weaknesses that deserve to be taken seriously. Previous studies have shown that the application of HMM in TCM needs to construct a corresponding estimation model for each tool wear state [8–15]. There are no unequivocal guidelines for the determination of the number of hidden states and possible observations in the HMM-based model. The number of hidden states and possible observations are only themselves within HMM. They have no specific relation with the actual tool wear states. So these parameters in HMM have no clear physical meaning. And inappropriate parameter selection will directly affect the prediction accuracy of the HMM-based classifier [17]. Consequently, the determination of the HMM-based optimal model usually requires a relatively long time. Generally, initialization of the parameters in each HMM-based model is

✉ Dongdong Kong
kodon007@163.com

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

randomly generated since no prior knowledge is available. The acquisition of an optimal model may need several training processes since the EM algorithm only guarantees local optimum. The above shortcomings make it difficult to establish optimal HMM-based models for TCM. Besides, previous studies on the application of HMM in TCM were centralized in two or three classification of tool wear states and so did not take full advantage of the superiority of HMM.

State duration distribution of the standard HMM is exponential duration density. This owes to the homogeneous Markov assumption, i.e., state of any moment depends only on the state of the previous moment. It indicates that the state duration probability in HMM shows an exponentially decline with time goes by. However, most of the practical applications do not satisfy the condition of exponential duration which will lead to inappropriate state duration modeling [18]. Hidden semi-Markov model (HSMM) is an extension of HMM and firstly proposed by Ferguson [19]. Yu [20] gives a complete overview of HSMM, including modeling, inference, estimation, and applications. "State duration" is a unique signature of HSMM within which each state can generate a series of observations. So the states in HSMM have no self-transition probabilities which are set to zero. This has greatly extended the application ranges of HSMM, such as handwritten word recognition [21], fault diagnosis [22, 23], and detection of reproductive status for dairy cows [24, 25]. In this paper, a HSMM-based monitoring system is utilized to estimate the tool wear state in milling process. Two groups of experiments are carried out to prove the effectiveness of the HSMM-based method. The cutting force signals are collected at set intervals under a complete tool wear process, i.e., initial wear, normal wear, severe wear, and breakage. Five types of time-domain features related to tool wear are extracted from the original cutting force signals. The HSMM-based method is utilized to construct the relationship between signal features and the corresponding tool wear state so as to realize on-line tool wear monitoring. It is noteworthy that the parameters in HSMM have clear physical meaning and the initial guess of parameters have certain rules to follow. Moreover, two kinds of state duration distribution, i.e., non-parametric and Gamma distribution are utilized together in the constructed HSMM-based tool wear monitoring system. The experimental results show that the HSMM with Gamma distribution has less iteration steps in the training process and is higher in prediction accuracy than other published methods. The convenience of model building, fast training speed, and high identification rate make HSMM a better means for tool wear monitoring. Besides, the identification of tool wear state can be implemented in less than 0.05 s by recurring to HSMM, which is very encouraging. This study provides a new method for tool wear monitoring in the field of intelligent manufacturing.

This paper proceeds as follows. Conceptual framework of HSMM and HSMM-based tool wear estimation model are introduced in Section 2. The experimental setup and data collection are presented in Section 3. Analysis of the HSMM-based monitoring system and experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2 Hidden semi-Markov model

Similar to other monitoring methods, the HSMM technique also consists of two parts: model training and state estimation. Detailed procedure is as follows.

2.1 HSMM components

HSMM contains the following basic contents:

1. State. Assuming the set of all the possible states is $S = \{q_1, q_2, \cdots, q_N\}$. N represents the number of possible states in HSMM. The corresponding state of an observed variable X_t at time t is defined as $S_t \in S$ which is unobservable.
2. Initial state probability vector $\boldsymbol{\pi} = [\pi_i]$.

$$\pi_i = P(S_1 = q_i), \quad i = 1, 2, \cdots, N. \tag{1}$$

where π_i represents the probability of being in state q_i at time $t = 1$ and satisfies the property $\sum_{i=1}^{N} \pi_i = 1$.

3. State transition probability matrix $\boldsymbol{A} = [a_{ij}]$.

$$a_{ij} = P\left(S_{t+1} = q_j | S_t = q_i\right), \quad i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, N. \tag{2}$$

where a_{ij} represents the probability of transition from state q_i at time t to state q_j at time $t+1$ and satisfies $\sum_{j=1}^{N} a_{ij} = 1$. Besides, $a_{ii} = 0$ since the self-transition probabilities are nonexistent in HSMM.

4. Observation distribution $\boldsymbol{B} = [b_i(\boldsymbol{x}_t)]$. The observed variable X_t at time t conditioned on the hidden state S_t is defined as observation distributions, which are either probability functions or probability densities [26]. In this study, multivariate normal distribution is utilized to fit the signal features observed in the machining process as following

$$\begin{aligned} b_i(\boldsymbol{x}_t) &= f(X_t = \boldsymbol{x}_t | S_t = q_i) \\ &= \frac{1}{(2\pi)^{k/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_i)\right), \end{aligned} \tag{3}$$

where k is the dimension of feature vectors. $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ represent the mean and covariance matrix of the observations, respectively.

5. State duration distribution $\boldsymbol{\Theta} = [d_i(u)]$. The probability density of a stochastic process spends the sojourn time or run-length u in a given state q_i is defined as state duration distribution. In HMM, the inherent duration probability in one state is $d_i(u) = a_{ii}^{u-1}(1-a_{ii})$, which is geometrically distributed and inappropriate for practical applications. State duration is modeled explicitly in HSMM which differs greatly from the HMM [26]. In this study, the non-parametric and Gamma [27, 28] distributions are utilized together to model the state duration.

For convenience, the HSMM can be briefly expressed by $\theta = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Theta})$. This set of model parameters θ needs to be estimated in the training process so as to generate accurate model.

2.2 HSMM model training

The signal features extracted from the milling process comprise the feature vectors, which further form the observation sequences and are utilized to estimate the model parameters θ. Supposing that the observed sequences O and the corresponding state sequences I are given by $O = (X_1, X_2, \cdots, X_T)$ and $I = (S_1, S_2, \cdots, S_T)$, respectively. In machining process, tool wear will go through four states in turn, i.e., initial wear, normal wear, severe wear, and breakage, and cannot go back. Tool breakage will not terminate and can continue indefinitely. Therefore, the time spent in the last state, i.e., tool breakage, can be considered as right-censoring. Guédon [29] proposed a new effective forward-backward algorithm to deal with the right-censoring in HSMM, which is utilized in this work. A brief introduction is as follows.

2.2.1 The EM algorithm

Complete data refers to $(O, I) = (X_1, X_2, \cdots, X_T, S_1, S_2, \cdots, S_T)$. The complete-data likelihood function of the HSMM can be expressed as

$$P(O, I|\theta) = \pi_1 d_1(u_1) \left\{ \prod_{r=2}^{N-1} a_{(r-1)r} d_r(u_r) \right\} a_{(N-1)N} D_N(u_N) \prod_{t=1}^{T} b_{S_t}(\boldsymbol{x}_t), \quad (4)$$

$$d_i(u) = P(S_{t+u+1} \neq i, S_{t+u-v} = i, v = 0, \cdots, u-2 | S_{t+1} = i, S_t \neq i). \quad (5)$$

where u_r represents the sojourn time spent in the state q_r. $D_j(u) = \sum_{v \geq u} d_j(v)$ is the survivor function of the run-length in the final visited state q_j. The utilization of this survivor function has two major benefits, i.e., improvement of parameter estimation and the guarantee of a more accurate prediction for the final visited state [25].

The model parameters θ can be reestimated by solving the maximum estimation of the incomplete-data log-likelihood function $L(\theta) = \log P(O|\theta)$. The EM algorithm [30] is utilized to deal with this maximization problem by iteration until convergence. The specific steps are as follows:

1. Firstly, initial value $\theta^{(0)}$ of the model parameters need to be determined.
2. Secondly, the expectation of the complete-data log-likelihood function needs to be calculated given the observed sequences O and the model parameters $\theta^{(l)}$,

$$Q\left(\theta, \theta^{(l)}\right) = E_I\left[\log P(O, I|\theta) | O, \theta^{(l)}\right] = \sum_I \log P(O, I|\theta) P\left(I | O, \theta^{(l)}\right). \quad (6)$$

3. Thirdly, the maximization of $Q(\theta, \theta^{(l)})$ is conducted so as to generate the reestimated parameters $\theta^{(l+1)}$,

$$\theta^{(l+1)} = \arg\max_\theta Q\left(\theta, \theta^{(l)}\right). \quad (7)$$

4. The two steps above are iterated as required until convergence,

$$\left\| L\left(\theta^{(l+1)}\right) - L\left(\theta^{(l)}\right) \right\| < \varepsilon. \quad (8)$$

2.2.2 Parameter reestimation

The reestimation for the model parameters $\theta = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Theta})$ of HSMM can be implemented by maximizing each subset of $Q(\theta, \theta^{(l)})$, respectively. Specific operation process of the EM algorithm for HSMM is summarized as following [24].

The E-step There are three items that need to be reestimated in the E-step for HSMM. The probability of the process being in state q_i at time t given the observed sequences O and the model parameters θ is denoted by

$$\gamma_t(i) = P(S_t = q_i | O, \theta). \quad (9)$$

The probability of the process transfers from state q_i at time t to state q_j at time $t+1$ is denoted by

$$\xi_t(i, j) = P\left(S_t = q_i, S_{t+1} = q_j | O, \theta\right). \quad (10)$$

State duration is modeled explicitly in HSMM. The probability of the process continues for u sojourn time in state q_i is denoted by

$$\begin{aligned} \eta_{iu} &= \eta'_{iu} + \eta^{`}_{iu}, \\ \eta'_{iu} &= P\left(S_{u+1} \neq q_i, S_{u-v} = q_i, v = 0, \cdots, u-1 | O, \theta\right), \\ \eta^{`}_{iu} &= \sum_{t=1}^{T-1} P(S_t \neq q_i, S_{t+u+1} \neq q_i, S_{t+u-v} = q_i, v = 0, \cdots, u-1 | O, \theta). \end{aligned} \quad (11)$$

The M-step The model parameters θ are reestimated in the M-step, respectively. The initial state probabilities are reestimated as

$$\pi_i^{(l+1)} = \gamma_1(i). \tag{12}$$

The state transition probabilities are reestimated as

$$a_{ij}^{(l+1)} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{i\neq j}\sum_{t=1}^{T-1}\xi_t(i,j)}. \tag{13}$$

In this study, multivariate normal distributions are utilized to fit the observed values as expressed by Eq. (3). The reestimation [31] for the parameters of multivariate normal distributions are given by

$$\boldsymbol{\mu}_i^{(l+1)} = \frac{\sum_{t=1}^{T}\gamma_t(i)\boldsymbol{x}_t}{\sum_{t=1}^{T}\gamma_t(i)}, \tag{14}$$

$$\boldsymbol{\Sigma}_i^{(l+1)} = \frac{\sum_{t=1}^{T}\gamma_t(i)(\boldsymbol{x}_t-\boldsymbol{\mu}_i)(\boldsymbol{x}_t-\boldsymbol{\mu}_i)^T}{\sum_{t=1}^{T}\gamma_t(i)}. \tag{15}$$

2.2.3 State duration distribution

The state duration density also needs to be reestimated in HSMM. Non-parametric and Gamma distributions are utilized together to model the state duration density explicitly in this study. Firstly, the non-parametric state duration distribution in the E-step can be given by

$$d_i(u) = \frac{\eta_{iu}}{\sum_v \eta_{iv}}. \tag{16}$$

The estimation items η_{iu} are the outputs of the E-step for duration distribution in state q_i. To obtain a concise HSMM, a parametric M-step in estimation procedure is utilized as a substitute for the non-parametric M-step of the EM algorithm [29]. In this way, the estimation items η_{iu} can be considered as the iteration terms produced by a given parametric duration distribution in state q_i so as to devise a parametric M-step. Secondly, the Gamma distribution with shape parameter α_i and scale parameter β_i, which is defined by $U_t|S_t = i \sim \Gamma(\alpha_i, \beta_i)$, is utilized to model the state duration distribution. Choi and Wette [28] presented the detailed derivation process for the reestimation of α_i and β_i. Supposing $\hat{\alpha}_i$ and $\hat{\beta}_i$ represent the reestimation of α_i and β_i that maximize the likelihood for the Gamma distribution, the solution can be obtained by solving

$$\begin{cases} \log(\hat{\alpha}_i)-\psi(\hat{\alpha}_i) = \log\left(\overline{u}_i\right)-\overline{w}, \\ \psi(\alpha_i) = \dfrac{d}{d(\alpha_i)}\Gamma(\alpha_i), \\ \overline{u}_i = \dfrac{\sum_u \eta_{iu}\cdot u}{\sum_u \eta_{iu}}, \\ \overline{w} = \dfrac{\sum_u \eta_{iu}\cdot \log(u)}{\sum_u \eta_{iu}}. \end{cases} \tag{17}$$

Equation (17) can be solved by utilizing the Newton-Raphson method. Once the shape parameter $\hat{\alpha}_i$ is generated, the scale parameter β_i can be estimated by $\hat{\beta}_i = \hat{\alpha}_i/\overline{u}_i$.

2.3 HSMM-based tool wear estimation

Once the training process is accomplished, the trained model parameters $\theta' = (\boldsymbol{\pi}', \boldsymbol{A}', \boldsymbol{B}', \boldsymbol{\Theta}')$ of HSMM that maximizes $L(\theta)$ are produced and can be employed to predict the state sequence of the new observed data. Strictly speaking, what we actually care about is the probability of being in state q_j at time T, which is defined by

$$\delta_t(j) = \max_{S_1,\cdots,S_{t-1}} P\left(S_{t+1}\neq q_j, S_t = q_j; S_1,\cdots,S_{t-1}; X_1,\cdots,X_{t-1}|\theta\right), \tag{18}$$

$$\delta_T(j) = \max_{S_1,\cdots,S_{T-1}} P\left(S_T = q_j; S_1,\cdots,S_{T-1}; X_1,\cdots,X_{T-1}|\theta\right). \tag{19}$$

The solution of the variable δ can be obtained via the Viterbi algorithm [29] as follows:

1. Initialization. The probability of being in state q_j at time $t = 1$ is given by

$$\delta_1(j) = b_j(\boldsymbol{x}_1)d_j(1)\pi_j. \tag{20}$$

2. Recursion. The probability of being in state q_j at time $t = 2, \cdots, T-1$ is given by

$$\begin{aligned} \delta_t(j) &= \max\left\{\vartheta'_t(j), \grave{\vartheta}_t(j)\right\}, \\ \vartheta'_t(j) &= b_j(\boldsymbol{x}_T)\left[\prod_{v=1}^{T-1} b_j(\boldsymbol{x}_{T-v})\right] D_j(T)\ \pi_j, \\ \grave{\vartheta}_t(j) &= b_j(\boldsymbol{x}_t)\max_{1\le u\le t-1}\left\{\left[\prod_{v=1}^{u-1} b_j(\boldsymbol{x}_{t-v})\right] d_j(u)\max_{i\neq j}\left[p_{ij}\delta_{t-u}(i)\right]\right\}. \end{aligned} \tag{21}$$

3. Termination. The probability of being in state q_j at time $t = T$ is given by

$$\begin{aligned} \delta_T(j) &= \max\left\{\vartheta'_T(j), \grave{\vartheta}_T(j)\right\}, \\ \vartheta'_T(j) &= b_j(\boldsymbol{x}_T)\left[\prod_{v=1}^{T-1} b_j(\boldsymbol{x}_{T-v})\right] D_j(T)\ \pi_j, \\ \grave{\vartheta}_T(j) &= b_j(\boldsymbol{x}_T)\max_{1\le u\le T-1}\left\{\left[\prod_{v=1}^{u-1} b_j(\boldsymbol{x}_{T-v})\right] D_j(u)\max_{i\neq j}\left[p_{ij}\delta_{T-u}(i)\right]\right\}. \end{aligned} \tag{22}$$

In this study, the optimal state corresponding to the observation sequence is considered as the actual state which is denoted by

$$q_j = arg\max_j \delta_T(j). \tag{23}$$

The observed sequence of length $T_0 (T_0 \le T)$ can also be considered as the right-censoring of the sojourn time. Consequently, the optimal state of the observation at time T_0

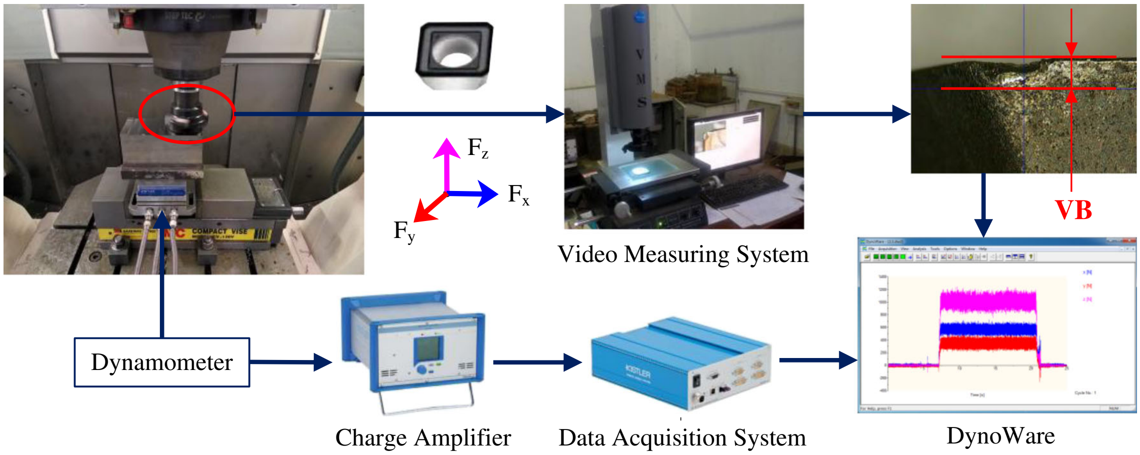


Fig. 1 The experimental setup

is identified as $arg\,\underset{j}{max}\,\delta_{T_0}(j)$ given the estimated model parameters θ'.

3 Experimental setup and data collection

The Experimental setup for tool wear estimation is illustrated in Fig. 1. The cutting tests are carried out on the Mikron UCP800Duro milling machine. The workpiece material is titanium alloy (Ti-6AI-4V (TC4)). Three inserts (Walter F2233.B.080.Z06.07) are installed on the cutter body (Walter SPMT1204AEN-WSP45) and symmetrically placed. The monitoring system consists of dynamometer (Kistler 9257A), charge amplifier (Kistler 5070A), data acquisition card (Kistler 5697A) and a laptop with DynoWare (Kistler 2825A). The DynoWare can visually display the variation of milling forces. The milling force signals in three dimensions (i.e., F_x, F_y, F_z) are collected at sampling rate of 5 kHz. The machining parameters are listed in Table 1. The tool flank wear is measured at set intervals in the machining process by utilizing the Video Measuring System (VMS-1510G (QIM1008)). The average of the three flank wear values are utilized as the standard to identify the tool wear state as shown in Table 2. The cutting tests will be continually conducted until tool breakage.

The inserts will go through four kinds of tool wear state in the machining process. The tool flank wear value is measured and photographed simultaneously. Tool flank wear appearances under four kinds of tool wear state are illustrated in Fig. 2. The effective features related to tool wear are extracted from the milling force signals so as to describe the tool wear state accurately. The mathematical representation of the extracted time-domain features is listed in Table 3. Each type of feature corresponds to three separate signal features (e.g., Mean_F_x, Mean_F_y, Mean_F_z). Besides, the signal features need to be normalized according to Eq. (24) so as to make them have the same order of magnitude. Meanwhile, the normalized signal features are also classified into four categories according to the tool wear state. Space distributions of these normalized features are illustrated in Fig. 3. It is obvious that the normalized features take on a certain degree of clustering property which can offer accurate information for tool wear estimation. The normalized signal features make up the feature vectors, which further form the observation sequence.

$$x' = \frac{x-\bar{x}}{\sigma_x} \tag{24}$$

where $\bar{x}$ is the mean value and σ_x is the standard deviation.

4 Experimental results and analysis

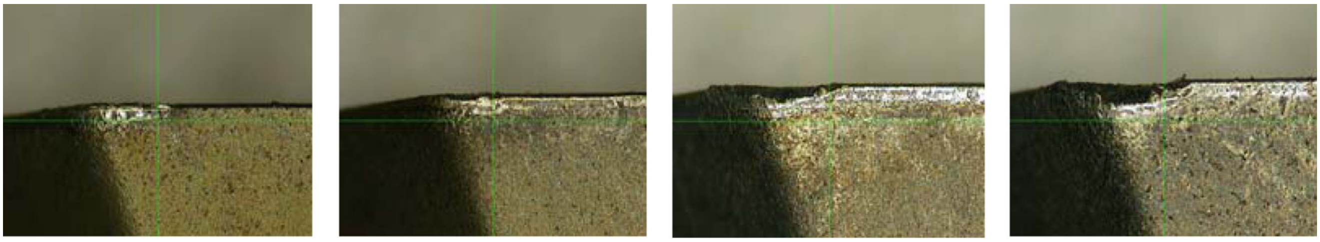
The utilization of HSMM for tool wear estimation is illustrated in this section. Two groups of experiments are conducted to

Table 1 The machining parameters

Test No.	Cutting speed (m/min)	Feed (mm/z)	Cutting depth (mm)	Cutting width(mm)
1	45	0.14	0.5	70
2	45	0.18	0.5	75

Table 2 Categories of tool wear state

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



Initial wear (0.138mm) Normal wear (0.213mm) Severe wear (0.2895mm) Breakage (0.3245mm)

Fig. 2 Tool flank wear appearances under four kinds of tool wear state (Test No. 1: insert 3)

prove the effectiveness of the HSMM-based monitoring system. In this study, each state in HSMM corresponds to one tool wear state. Thus, the physical meaning of the states in HSMM is clear ($N = 4$): State1 − initial wear, State2 − normal wear, State3 − severe wear, and State4 − breakage. There are altogether 15 signal features as listed in Table 3 that constitute the feature vector which further makes up the observed sequence. The observation sequences that collected in each tool wear state at set intervals are randomly divided into training and testing samples, which have the same size and do not contain each other.

The tool wear in the machining process begins with the initial wear, undergoes normal wear and severe wear and terminates at tool breakage. Thus, the "left-right" topology is selected for the determination of the HSMM structure since the tool wear process cannot go back in theory. The initialization of the model parameters $\theta^{(0)} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Theta})$ needs to be carried out before the training process. Parameter initialization is of vital importance for the model fitness since the EM algorithm only guarantee local optimal solution. The initialization of the model parameters $\theta^{(0)}$ is as follows.

The initial state probability vector should be $\boldsymbol{\pi} = [1\ 0\ 0\ 0]$ since the tool wear starts from initial wear. Each state in HSMM can generate a series of observations and state self-transition is non-existent. Therefore, the main diagonal elements in $\boldsymbol{A}$ should be set to zero. It is noteworthy that the initial wear directly transfers to severe wear or tool breakage in $\boldsymbol{A}$ is also permitted since tool fracture may occur suddenly in machining process even if the probability is very small. Besides, the turning back of tool wear in $\boldsymbol{A}$ is also permitted since the emergence of built-up edge may generate some observations belong to initial wear or normal wear, although the probability is comparatively small. The state transition probability matrix is set as

$$\boldsymbol{A} = \left[a_{ij}\right] = \begin{array}{c} \\ \left[\begin{array}{cccc} \text{state1} & \text{state2} & \text{state3} & \text{state4} \\ 0 & 0.90 & 0.05 & 0.05 \\ 0.05 & 0 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0 & 0.90 \\ 1/3 & 1/3 & 1/3 & 0 \end{array}\right] \end{array}$$

As for the parameters of observation distribution $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i (i = 1, 2, 3, 4)$, they are initialized according to the corresponding observations of training data in each state. The mean of the feature vectors in the observation under four kinds of tool wear state are listed in Table 4. The covariance matrix $\boldsymbol{\Sigma}_1$ of the observation under the initial wear state is listed in Table 5. $\boldsymbol{\Sigma}_i (i = 2, 3, 4)$ are no longer shown due to space limitations.

The unique feature of HSMM lies in the state duration which makes it greatly different from a standard HMM. State duration is modeled explicitly in HSMM. In this study, the non-parametric and Gamma distributions are utilized together to model the state duration. The parameters in the non-parametric distribution can be denoted by a matrix $[d_i(u)]_{M \times N}$ in which each element (u, i) corresponds to a probability $d_i(u)$. M represents the maximum sojourn time of one state in the machining process. N represents the number of tool wear states. A uniform distribution which covers the starting and ending time of each state is a good initial guess for the non-parametric distribution when there is no heuristic knowledge.

Table 3 Mathematical representation of the time-domain features

Type	Signal features	Mathematical expression
Mean	Mean_F$_x$, Mean_F$_y$, Mean_F$_z$	$\mu = E(\lvert x_i \rvert)$
Maximum (max)	Max_F$_x$, Max _F$_y$, Max _F$_z$	$x_{\text{Max}} = \max(\lvert x_i \rvert)$
Peak to valley (PV)	PV_F$_x$, PV_F$_y$, PV_F$_z$	$x_{\text{PV}} = x_{\text{Max}} - x_{\text{Min}}$
Root mean square (RMS)	RMS_F$_x$, RMS_F$_y$, RMS_F$_z$	$x_{\text{RMS}} = \{E(x_i^2)\}^{1/2}$
Standard deviation (Std)	Std_F$_x$, Std_F$_y$, Std_F$_z$	$x_{\text{Std}} = \{E[(\lvert x_i \rvert - \mu)^2]\}^{1/2}$

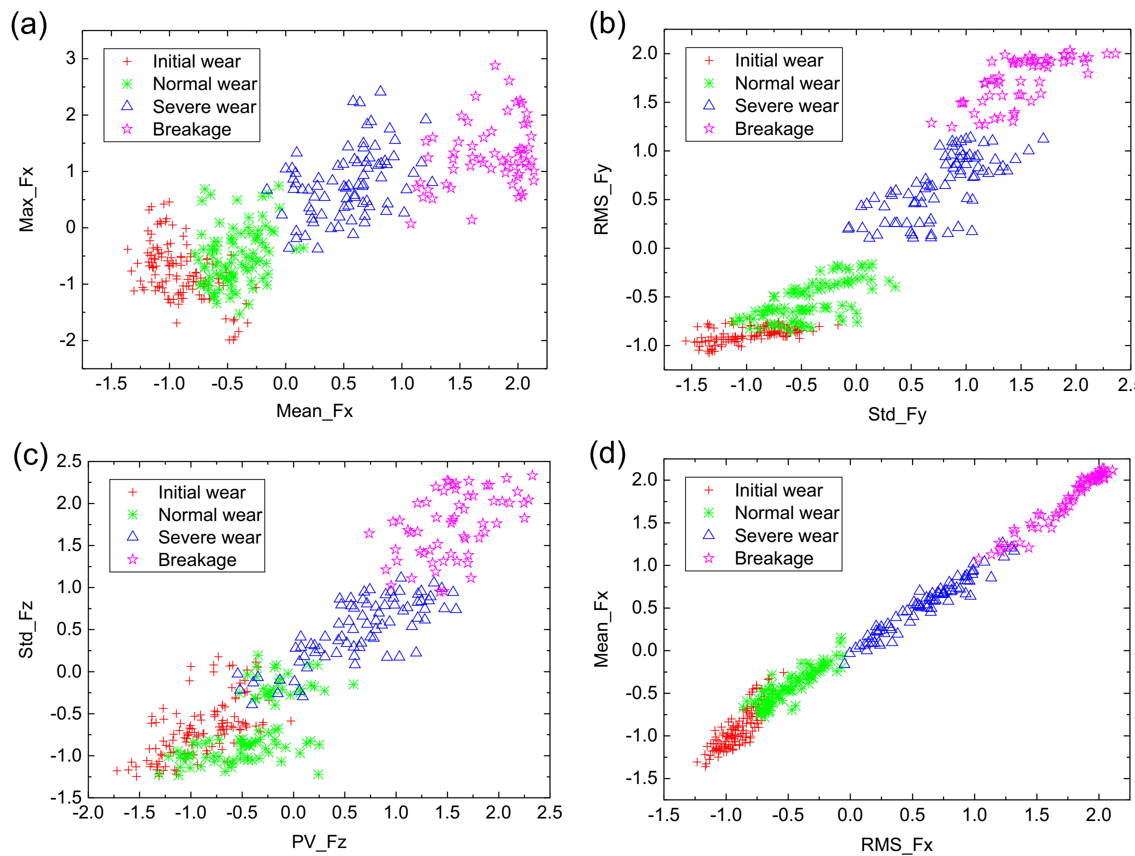


Fig. 3 Space distribution of the normalized features under four kinds of tool wear state (Test No. 1)

Supposing that the maximum state sojourn time $M = 200$, the initialization for $[d_i(u)]_{M \times N}$ is given by

$$\boldsymbol{\Theta} = [d_i(u)]_{M \times N} = \begin{array}{c} \\ \left[\begin{array}{cccc} \text{state1} & \text{state2} & \text{state3} & \text{state4} \\ 0.005 & 0.005 & 0.005 & 0.005 \\ 0.005 & 0.005 & 0.005 & 0.005 \\ \vdots & \vdots & \vdots & \vdots \\ 0.005 & 0.005 & 0.005 & 0.005 \end{array} \right]_{200 \times 4} \end{array}$$

As mentioned in Section 2.2.3, a parametric M-step in estimation procedure is utilized as a substitute for the non-parametric M-step of the EM algorithm. To obtain a concise HSMM, the uniform distribution $d_i(u)$ with a reasonable range of values is utilized as the initial guess for the Gamma distribution since reasonable starting parameters are not available [25].

The training for the HSMM model can be implemented when the initialization of the model parameters $\theta^{(0)} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Theta})$ is finished. The training process is implemented by utilizing the "mhsmm" package [32]. The iteration stop condition ε is set as 0.0001. The training procedures for HSMM under the Gamma distribution are illustrated in Fig. 4. The convergence of the EM algorithm for HSMM under the Gamma distribution requires only 29 iterations. The training time of HSMM under the Gamma distribution for Test No. 1 and Test No. 2 are 0.386 and 0.356 s, respectively. The estimated initial state probability vector and state transition probability matrix for the two groups of experiments are as follows:

$$\boldsymbol{\pi}' = [\pi_i] = [1\ 0\ 0\ 0],$$

$$\boldsymbol{A}' = [a_{ij}] = \left[\begin{array}{cccc} \text{state1} & \text{state2} & \text{state3} & \text{state4} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

It is obvious that the training results basically agree with the changing process of tool wear states, which also proves the feasibility of HSMM. The estimated model parameters of the Gamma distribution in HSMM are listed in Table 6. This parametric state duration distribution makes the HSMM more concise. The estimated parameters of the observation distribution $\boldsymbol{\mu}_i'$ and $\boldsymbol{\Sigma}_i'$ are almost unchanged in comparison with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ since the initial values are calculated from the corresponding observations.

The trained HSMM model θ' that maximizes $L(\theta)$ is generated when the training process is accomplished and can be utilized to predict tool wear status. Prediction of tool wear

Table 4 The mean of the feature vectors in the observation under four kinds of tool wear state (Test No. 1)

μ_i	Mean_F_x	Mean_F_y	Mean_F_z	Max_F_x	Max_F_y	Max_F_z	PV_F_x	PV_F_y	PV_F_z	RMS_F_x	RMS_F_y	RMS_F_z	Std_F_x	Std_F_y	Std_F_z
State1 − μ_1	−0.9148	−0.8769	−0.8735	−0.7882	−0.8828	−0.9172	−0.7882	−0.8893	−0.8961	−0.9052	−0.8834	−0.8658	−0.8443	−0.9449	−0.6885
State2 − μ_2	−0.4249	−0.5517	−0.4887	−0.5084	−0.4592	−0.4218	−0.5085	−0.4536	−0.4162	−0.4860	−0.5451	−0.4999	−0.5895	−0.4320	−0.6831
State3 − μ_3	0.5343	0.6311	0.5139	0.7872	0.6787	0.5603	0.7873	0.7094	0.6440	0.6022	0.6383	0.5123	0.7163	0.7677	0.4943
State4 − μ_4	1.7545	1.7400	1.7948	1.2894	1.5495	1.7199	1.2893	1.5131	1.5578	1.7365	1.7335	1.7984	1.6177	1.5143	1.7424

Table 5 The covariance matrix $\boldsymbol{\Sigma}_1$ of the observations under the initial wear state (Test No. 1)

Σ_1	Mean_F_x	Mean_F_y	Mean_F_z	Max_F_x	Max_F_y	Max_F_z	PV_F_x	PV_F_y	PV_F_z	RMS_F_x	RMS_F_y	RMS_F_z	Std_F_x	Std_F_y	Std_F_z
Mean_F_x	0.0574	0.0086	0.0276	−0.0625	−0.0039	0.0323	−0.0625	−0.0040	0.0208	0.0281	0.0083	0.0267	−0.0361	0.0060	0.0119
Mean_F_y	0.0086	0.0057	0.0057	−0.0065	0.0036	0.0046	−0.0065	0.0037	0.0059	0.0057	0.0055	0.0055	−0.0014	0.0049	0.0027
Mean_F_z	0.0276	0.0057	0.0236	−0.0404	−0.0044	0.0250	−0.0405	−0.0045	0.0156	0.0120	0.0056	0.0232	−0.0222	0.0049	0.0175
Max_F_x	−0.0625	−0.0065	−0.0404	0.2538	0.0464	−0.0310	0.2539	0.0478	0.0158	−0.0073	−0.0015	−0.0364	0.1104	0.0481	0.0357
Max_F_y	−0.0039	0.0036	−0.0044	0.0464	0.0985	0.0043	0.0464	0.1015	0.0294	0.0122	0.0091	−0.0014	0.0457	0.0637	0.0541
Max_F_z	0.0323	0.0046	0.0250	−0.0310	0.0043	0.0513	−0.0311	0.0044	0.0629	0.0162	0.0057	0.0259	−0.0190	0.0172	0.0460
PV_F_x	−0.0625	−0.0065	−0.0405	0.2539	0.0464	−0.0311	0.2540	0.0478	0.0158	−0.0073	−0.0016	−0.0365	0.1104	0.0480	0.0356
PV_F_y	−0.0040	0.0037	−0.0045	0.0478	0.1015	0.0044	0.0478	0.1046	0.0303	0.0125	0.0093	−0.0015	0.0471	0.0656	0.0557
PV_F_z	0.0208	0.0059	0.0156	0.0158	0.0294	0.0629	0.0158	0.0303	0.1247	0.0172	0.0085	0.0185	0.0087	0.0352	0.0733
RMS_F_x	0.0281	0.0057	0.0120	−0.0073	0.0122	0.0162	−0.0073	0.0125	0.0172	0.0197	0.0068	0.0123	0.0005	0.0185	0.0188
RMS_F_y	0.0083	0.0055	0.0056	−0.0015	0.0091	0.0057	−0.0016	0.0093	0.0085	0.0068	0.0062	0.0057	0.0027	0.0130	0.0090
RMS_F_z	0.0267	0.0055	0.0232	−0.0364	−0.0014	0.0259	−0.0365	−0.0015	0.0185	0.0123	0.0057	0.0231	−0.0193	0.0083	0.0224
Std_F_x	−0.0361	−0.0014	−0.0222	0.1104	0.0457	−0.0190	0.1104	0.0471	0.0087	0.0005	0.0027	−0.0193	0.0783	0.0442	0.0327
Std_F_y	0.0060	0.0049	0.0049	0.0481	0.0637	0.0172	0.0480	0.0656	0.0352	0.0185	0.0130	0.0083	0.0442	0.0932	0.0719
Std_F_z	0.0119	0.0027	0.0175	0.0357	0.0541	0.0460	0.0356	0.0557	0.0733	0.0188	0.0090	0.0224	0.0327	0.0719	0.1168

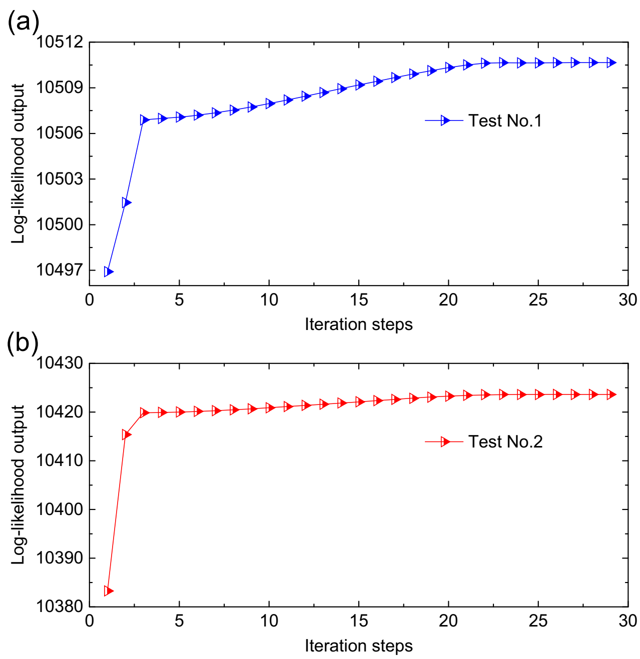


Fig. 4 The training procedures for HSMM under the Gamma distribution ($M = 200$)

states in machining process is carried out based on the Viterbi algorithm as introduced in Section 2.3. The observed sequence of length $T_0(T_0 \leq T)$ can also be considered as right-censoring. Consequently, the optimal state of the observation at time $T_0(T_0 \leq T)$ is identified as $\arg\max_j \delta_{T_0}(j)$ given the estimated model parameters θ'. Monitoring process of the tool wear state by utilizing HSMM under the Gamma distribution are illustrated in Fig. 5. It is clear that the misclassified points exist only in the transitions between consecutive states which guarantees the stability of the monitoring process. The prediction accuracy for four kinds of tool wear state in Test No. 1 and Test No. 2 reach up to 99.43 and 98.39%, respectively. Besides, the testing time of HSMM under the Gamma distribution for Test No. 1 and Test No. 2 are 0.045 and 0.046 s, respectively. High identification rate and fast identification speed make the HSMM especially propitious to tool wear monitoring in industrial environment.

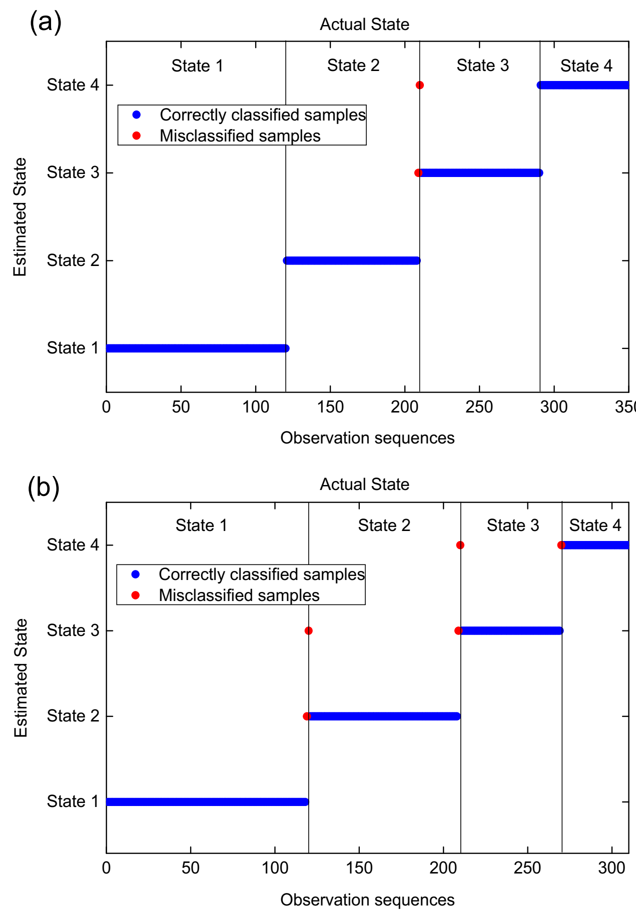


Fig. 5 Monitoring process of the tool wear state by utilizing HSMM under the Gamma distribution: (**a**) Test No. 1 and (**b**) Test No. 2

4.1 Comparison with other methods

To verify the reliability and effectiveness of the HSMM-based monitoring system, some other published methods are tested using the same set of training and testing data. C-support vector classification (C-SVC) [33] and Back-Propagation neural network (BPNN) [7] are utilized to realize multi-class classification of tool wear states. In C-SVC, the One-Vs-One voting method is employed to achieve multi-classification. Besides, radial basis kernel function (RBF) is adopted so as to improve the model accuracy. The model parameters

Table 6 The estimated model parameters of the Gamma distribution in HSMM

	$(\hat{\alpha}_i, \hat{\beta}_i)$	State1	State2	State3	State4
Test No. 1	Shape : $\hat{\alpha}_i$	55,219.21	32,393.84	25,574.29	14,398.32
	Scale : $\hat{\beta}_i$	0.002154727	0.002778300	0.003128181	0.004167146
Test No. 2	Shape : $\hat{\alpha}_i$	55,219.886	32,395.059	13,890.699	6710.162
	Scale : $\hat{\beta}_i$	0.002154700	0.002778198	0.004247315	0.006110460

Table 7 Tool wear states and the corresponding codes

States	State1	State2	State3	State4
Codes	0001	0010	0100	1000

$(C, \gamma) \in [2^{-10}, 2^{10}] \times [2^{-5}, 2^5]$ need to be optimized and are determined via the grid-search method and 10-fold cross-validation. In BPNN, the four kinds of tool wear state need to be encoded as listed in Table 7. The linear and sigmoid activation functions are selected as the hidden layer and output layer, respectively. The hidden layer has 25 nodes which are determined by trial and error. The learning rate is set to 0.05. The maximum number of iterations is set to 10,000. The identification rate for four kinds of tool wear state by utilizing HSMM, C-SVC, and BPNN are illustrated in Fig. 6. The predicted results reveal that the HSMM-based model has better prediction accuracy than C-SVC and BPNN.

The advantages of the proposed method have been apparent in comparison with the above mathods: (a) higher prediction accuracy, (b) the model structure is easy to be determined, and (c) parameter optimization is not required. It can be concluded that the HSMM-based monitoring system is ideally qualified for tool wear estimation.

5 Conclusions

This paper presents a new tool wear monitoring system based on hidden semi-Markov model (HSMM). The Gamma distribution is utilized to model the state duration which is initialized by a non-parametric distribution. Milling experiments are carried out to prove the effectiveness of the proposed method. Five types of time-domain feature related to tool wear are extracted from the milling force signals and utilized to realize tool wear estimation. The experimental results reveal that the prediction accuracy of HSMM under the Gamma distribution exceeds 98%. Comparison results reveal that the proposed method has relatively higher prediction accuracy than C-SVC and BPNN. High identification rate and fast identification speed (less than 0.05 s) make the proposed method especially propitious to tool wear monitoring. Besides, the model structure of HSMM is more easily determined and parameter optimization is not required, which greatly reduce the time-cost for practical application in industrial environment.

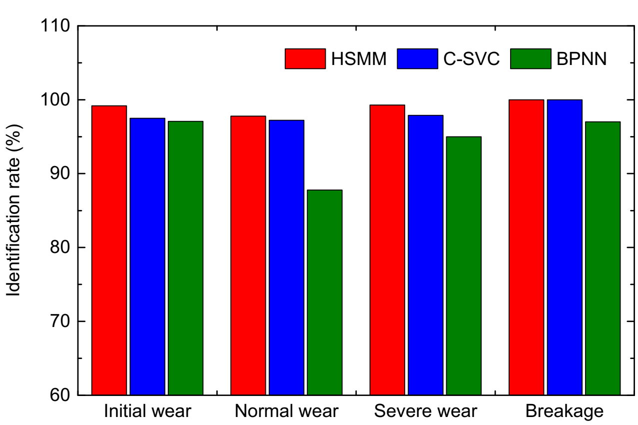


Fig. 6 The identification rate for four kinds of tool wear state by utilizing HSMM, C-SVC, and BPNN

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

References

1. Tönshoff HK, Wulfsberg JP, Kals HJJ, König W, van Luttervelt CA (1988) Developments and trends in monitoring and control of machining processes. CIRP Ann Manuf Technol 37(2):611–622
2. Kramer BM, Von Turkovich BF (1986) A comprehensive tool wear model. CIRP Ann Manuf Technol 35(1):67–70
3. Mehrabi MG, Ulsoy AG, Koren Y (2000) Reconfigurable manufacturing systems: key to future manufacturing. J Intell Manuf 11:403–419
4. 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
5. 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
6. Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. Mech Syst Signal Process 21: 1799–1814
7. 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
8. Wang LT, Mehrabi MG, Elijah KA (2002) Hidden Markov model-based tool wear monitoring in turning. J Manuf Sci Eng 124(3): 652–658
9. 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
10. 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
11. 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
12. 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
13. 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
14. 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
15. Baruah P, Chinnam RB (2005) HMMs for diagnostics and prognostics in machining processes. Int J Prod Res 43(6):1275–1293

16. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
17. Wang GF, Feng XL (2013) Tool wear state recognition based on linear chain conditional random field model. Eng Appl Artif Intell 26:1421–1427
18. Russell MJ, Moore RK (1985) Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition. Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing 10:5–8
19. J.D. Ferguson (1980) Variable duration models for speech. Proceedings of the Symposium on the Application of Hidden Markov Models to Text & Speech, Princeton, New Jersey, Oct 1980, pp 143–179
20. Yu SZ (2010) Hidden semi-Markov models. Artif Intell 174:215–243
21. A. Kundu, Y. He and M.Y. Chen (1997) Efficient utilization of variable duration information in HMM based HWR systems. Proceedings of International Conference on Image Processing, 26–29 Oct 1997, pp 304–307
22. Dong M, He D (2007) A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. Mech Syst Signal Process 21:2248–2266
23. Dong M, He D (2007) Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. Eur J Oper Res 178:858–878
24. O'Connell J, Tøgersen FA, Friggens NC, Løvendahl P, Højsgaard S (2010) Combining cattle activity and progesterone measurements using hidden semi-Markov models. Journal of Agricultural Biological and Environmental Statistics 16(1):1–16
25. O'Connell J, Højsgaard S (2011) Hidden semi Markov models for multiple observation sequences: the mhsmm package for R. J Stat Softw 39(4):1–22
26. Bulla J, Bulla I, Nenadić O (2010) hsmm—an R package for analyzing hidden semi-Markov models. Computational Statistics and Data Analysis 54:611–619
27. Levinson SE (1986) Continuously variable duration hidden Markov models for automatic speech recognition. Comput Speech Lang 1(1):29–45
28. Choi SC, Wette R (1969) Maximum likelihood estimation of the parameters of the Gamma distribution and their bias. Technometrics 11(4):683–690
29. Guédon Y (2003) Estimating hidden semi-Markov chains from discrete sequences. J Comput Graph Stat 12(3):604–639
30. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1):1–38
31. J.A. Bilmes (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
32. mhsmm: Inference for Hidden Markov and Semi-Markov Models. <https://cran.r-project.org/web/packages/mhsmm/index.html>
33. Bordoloi DJ, Tiwari R (2014) Optimum multi-fault classification of gears with integration of evolutionary and SVM algorithms. Mech Mach Theory 73:49–60