This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

Diagnosis and Prognosis of Degradation Process via Hidden Semi-Markov Model

Tongshun Liu , Kunpeng Zhu *Member IEEE*, Liangcai Zeng

***Abstract*—The intelligent estimation of degradation state and prediction of remaining useful life (RUL) are important to the maintenance of industrial equipment. In this study, the degradation process of equipment is modeled as an improved hidden semi-Markov model (HSMM), in which the dependence of durations of adjacent degradation states is described and modeled in the HSMM. To avoid underflow problem in computing the forward and backward variables, a modified forward-backward algorithm is proposed in the HSMM. Based on the improved algorithm, on-line estimation of degradation state and the distribution of remaining useful life (RUL) can be obtained. Case studies on tool wearing diagnosis and prognosis have verified the effectiveness of this model.**

***Index Terms*—Degradation process, health monitoring, remaining useful life, Hidden semi-Markov, forward-backward algorithm**

I. Introduction

Degradation of industrial equipment such as wearing and erosion is a common issue in manufacturing process, and it restricts productivity greatly. Efficient approach that can monitor the degradation conditions and predict remaining useful life of equipment with degradation process is necessary for improving producing quantity and saving cost [1][2]. In the machinery condition monitoring and fault diagnosis, model-based approaches [3] and data-based approaches [4] are usually adopted to estimate the health condition and RUL. As a kind of model-based approach, stochastic processes have been widely studied to model degradation [5]-[8]. In [5], degradation process was modeled as a Wiener process whose parameters were constantly updated by recursive filtering algorithm and expectation-maximization (EM) algorithm. Braglia et.al [6] utilized stochastic diffusion process to represent the wearing process of cutting tool. The model-based approach requires that the current degradation state could be measured directly when applied for prognostics. Hence, it is hard to model and update in many applications in which the degradation level could not be explicitly acquired.

Resorting to statistical and machine learning approaches, data-based approaches utilize observational signals to infer the health state and RUL of equipment. The first step of data-based approach is extracting features that could indicate the degradation state from sensory signals [9-10]. In [9], wavelet energy spectrum of vibration signal was adopted for bearing fault detection. Zhu et.al [10] found that the distribution of Holder exponent of cutting force signal indicated tool wearing process. To get discriminated features from original collection of features, techniques such as Fisher discriminant analysis [11], sparse decomposition [12] and principle component analysis [13] were usually applied. Following the features extraction, intelligent approach is then applied for diagnosis and prognosis, which is either pure data-based or model–data integrated methodologies [4]. In pure data-based approaches such as SVM [14] and neural networks [15], only the observation features are utilized to infer the hidden degradation state and the latent evolution law of degradation is not taken into account. However, in real applications, it is almost impossible to extract distinct features for different degradation states. The feature spaces of different degradation states may include outliers and overlaps and the recognition error will be high when these features occur. In model-data integrated approaches such as Kalman filtering [16] [17] and HMM [18], both the observation model and system model are built. The observation model represents the relationship between data features and states and the system model reflects the latent evolution law of hidden states. In this way, it utilizes both side information and provides better diagnostic capabilities. Kalman filtering always needs to assign analytical form for state and observation equations. This makes Kalman filtering less convenient than HMM in which the inherent structure could be learned based on experimental data alone. Moreover, Kalman filtering could not handle non-Gaussian noise well. By modeling the non-Gaussian noise as a non-Gaussian transition probability, HMM is more powerful to deal with it.

For these advantages compared to pure data-based approaches and Kalman filtering, HMM has been widely used in health monitoring. According to the topology of HMM, existing HMM approaches could be divided into three categories: hidden Markov models (HMMs) [19]-[23], single HMM [24]-[29], and two-layer HMM [30] [31]. Approaches via HMMs assign one hidden Markov model for each health condition. The health condition which corresponds to one HMM with maximum likelihood for observation sequence is seen as estimated health condition. Athanasopoulou et.al [19] proposed an on-line recursive algorithm to choose the most likely HMM for given erroneous observations from multi HMMs corresponding to different failure modes. Approaches via single HMM build the entire varying process of health condition as a hidden Markov process. Given observations, the condition that corresponds to the state with maximum posterior

This work was supported in part by the CAS 100 Talents Program, Chinese Academy of Sciences, and in part by the National Natural Science Foundation of China under Grant 51475443. (Corresponding author: Kunpeng Zhu.)

T. S. Liu is with the Department of Automation, University of Science and Technology of China, Hefei, 230026, China (e-mail: zkdlts@mail.ustc.edu.cn).

K. P. Zhu is with the Institute of Advanced Manufacturing Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Huihong Building, Changwu Middle Road 801, Changzhou 213164, Jiangsu, China (e-mail: kunpengz@hotmail.com).

L. C. Zeng is with the School of Machinery and Automation, Wuhan University of Science and Technology, #947 Heping Avenue, Qingshan District, Wuhan 430081, Hubei, China (e-mail: zengliangcai@wust.edu.cn).

probability is seen as estimated health condition. In [24], fault propagation process was modeled as a high-order hidden Markov process. Then high-order particle filtering was utilized to obtain the posterior estimate of state. Approaches via two-layer HMM not only include the global Markov transition of state but also include the Markov transition of sub-state. In [30], two-layer hierarchical HMM was adopted to represent the two kinds of Markov transition relationships.

In spite of the wide applications of HMMs in machinery health monitoring, the time-invariant one-step transition probability and geometric distribution of duration make the HMM not match real degradation process well. For instance, unlike the conventional HMM assumption, the probability that the wear state of a tool changes would mostly decrease as the cutting time increases. As a generalized HMM, the hidden semi-Markov model (HSMM) incorporates self-transition probability into distribution of duration of state. Different forms of distribution of duration correspond to different forms of time-variant self-transition probability. To model inhomogeneous Markov process, HSMM is more flexible and reasonable than HMM. For this reason, HSMM with explicit duration is widely used in health monitoring domain [32]-[35]. Explicit duration HSMM assigns an explicit distribution of duration for each state and the distribution of duration is only determined by corresponding state. The time variant one-step transition probability of explicit duration HSMM is dependent on the current state and the lasted time of current state. Geramifardet.al [34] showed that the diagnostic result via a single explicit duration HSMM is more accurate than that via a single HMM in [27]. Another special form of HSMM commonly used in health monitoring is variable transition HSMM [36] [37]. Instead of assigning explicit distribution of durations, variable transition HSMM assigns explicit time-variant form for one-step transition probability. Like explicit duration HSMM, the one-step transition probability of variable transition HSMM is also dependent on current state and corresponding lasted time. In [36], Peng et.al used variable transition HSMM with analytical one-step transition probability to estimate RUL. By taking the duration-dependent transition of degradation state into account, Wang et.al [37] developed a forward-backward algorithm to estimate the remaining useful life of bearing. Of all the above studies focusing on applying HSMM to health monitoring, the dependence between durations of different states were not taken into account. However, there exists dependence between durations of different health states in real case. And for degradation process, the dependent relationship between durations of adjacent states is often remarkable. For example, a long duration in fresh tool state may lead to a long duration in medium wear state. Therefore, including the dependent relationship between durations into the HSMM for degradation process is more reasonable.

Based on such consideration, this study develops an improved HSMM model that includes sequential transition of state and dependent durations of adjacent degradation states to describe the degradation process. In this model, the one-step transition probability of degradation process is not only dependent on the current state and lasted time but also dependent on the duration of preceded state. To avoid the underflow problem in the estimating of state and RUL, a modified forward-backward algorithm is proposed. And based on the modified forward-backward algorithm, on-line health monitoring and RUL estimation are then conducted conveniently. In this study, different from the prognosis methods via HSMM in [32] [33] [37], the distribution of RUL rather than limited statistics of RUL is derived to provide more comprehensive information. The paper is organized as follows. In section II, hidden semi-Markov model for degradation process is built. On-line health monitoring via modified forward-backward algorithm for proposed HSMM is given in section III. To validate the improvement of proposed HSMM, in section IV, numerical simulation via tool wear monitoring experiment is given, and concluded in section V.

II. Hidden Semi-Markov Model for Degradation Process Modeling

The hidden semi-Markov model [38] is a generalized hidden Markov model in which the transition from state and duration in a time segment to the state and duration in next segment is considered. Different from the HMM, the transition in HSMM is a transition between 2-D vectors (state and duration) in adjacent time segments. State variable at time t is denoted by q_t and observation variable is o_t. A hidden semi-Markov model is described by following three elements:

Transition probability:

$$a_{(i,d)(j,d^{'})} \triangleq P\left\{q_{|t+1:t+d^{'}|} = s_j \mid q_{|t-d+1:t|} = s_i\right\} \tag{1}$$

Observation probability:

$$c_i(v_j) \triangleq P\left\{o_t = v_j \mid q_t = s_i\right\} \tag{2}$$

Initial distribution:

$$\pi_{i,d} \triangleq P\left\{q_{|t-d+1:t|} = s_i\right\} \tag{3}$$

where s_i is ith state and v_j is jth observation. The set of model parameters is defined as:

$$\lambda = \left\{a_{(i,d)(j,d^{'})}, c_i(v_j), \pi_{j,d}\right\} \tag{4}$$

To model degradation process via HSMM, degradation process is divided into N discrete states $\left\{s_i \mid i = 1...N\right\}$ according to underlying degradation level. State s_1 represents the initial state with slight degradation level and state s_N represents final state with severe degradation level. By assuming that the degradation state changes sequentially from initial state to final state and the durations of adjacent states are dependent, the transition probability of HSMM for degradation process could be simplified as:

$$a_{d^{'}(i,d)} \triangleq P\left\{q_{|t+1:t+d|} = s_i \mid q_{|t-d^{'}+1:t|} = s_{i-1}\right\} \tag{5}$$

Namely, the transition probability represents the conditional probability of duration of latter state on condition that the duration of preceded state is given. The degradation process is assumed to be irreversible and it will always stays at the final state as soon as it reaches to the final state. So the duration of the final state is infinite and there is:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

$$a_{d^{'}(N,+\infty)} = 1 \quad \forall d^{'} \tag{6}$$

Initial distribution of HSMM for degradation process is defined as the distribution of duration of initial state:

$$p_1(d) \triangleq P\left\{q_{|1:d|} = s_1\right\} \tag{7}$$

Generally, the observation is continuous feature vector extracted from physical signal. To process the high-demission observation, two approaches are usually adopted. One approach is using continuous distribution with high demission such as multinormal distribution to descript the relationship between observation and state. The other approach is quantizing the observational feature via vector quantization (VQ) technique [39], and then the relationship between observation and state is modeled by discrete distribution function. In our paper, without assigning explicit distribution form for feature vector, the second approach is adopted. The M possible discrete observation produced by degradation process is denoted by $\{v_j \mid j = 1...M\}$ and observation probability is denoted by $c_i(v_j)$. A special HSMM for degradation process is defined as:

$$\lambda_d = \left\{a_{d^{'}(j,d)}, c_i(v_j), p_1(d)\right\} \tag{8}$$

The parameters of HSMM could be estimated with experimental data. When the observations and states are both recorded completely in experimental process, maximum likelihood method could be used directly to estimate the parameters. When the state data cannot be recorded in experiments, with lifetime observation sequence of a degradation process only, the off-line training of HSMM for degradation process should resort to EM algorithm. Based on the trained HSMM, with current and historical observations, on-line estimation of current degradation state and remaining useful life is then conducted. The diagram of off-line training and on-line health monitoring via HSMM is given in Figure 1.

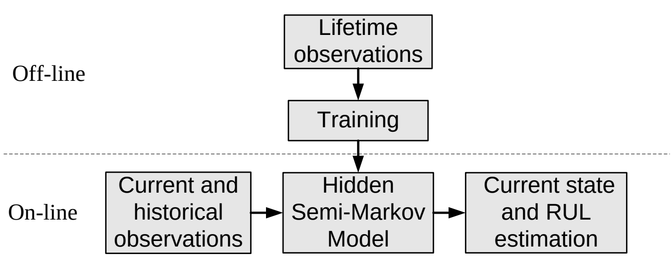


Fig.1. Diagram of off-line training and on-line health monitoring via HSMM

III. On-line Health Monitoring via HSMM

In the theory of HSMM, forward and backward variables are two fundamental variables. Almost all of the posterior probabilities needed in inference could be represented by the two variables. Traditional forward-backward algorithm with forward (backward) variable as joint probability of observation sequence encounters severe underflow problem. In this section, to conquer underflow problem in HSMM for degradation process, we propose a modified forward-backward algorithm. Similar to the forward-backward algorithm which is proposed for explicit duration HMM in [40], the modified forward-backward algorithm defines forward variable as a posterior probability of observation sequence. According to the modified forward-backward algorithm, the posterior probability of degradation state and RUL could be calculated conveniently.

A. *Modified Forward-Backward Algorithm*

The observation sequence is denoted by $o_1 o_2 ... o_{T_l}$. For $t \le T_l$, the formal forward and backward variables [38] are defined as:

$$\alpha_t(i,d) \triangleq P\left\{q_{|t-d+1:t|} = s_i, o_{1:t} \mid \lambda_d\right\} \tag{9}$$

$$\beta_t(i,d) \triangleq P\left\{q_{|t-d+1:t|} = s_i, o_{t+1:T_l} \mid \lambda_d\right\} \tag{10}$$

where $q_{|t-d+1:t|} = s_i$ means that the state becomes s_i at time $t-d+1$ and changes to another state at time $t+1$. It could be found that the forward variable is generally very small when the time t lasts long and observation features are high. When the forward variables at time t are less than the decimal of the computation system, they will be set as zero and the underflow phenomenon occurs. Then the forward variable cannot be used in the following inference or estimation process. The underflow phenomenon also occurs often in the computation of backward variables (10) when the time t is far from T_l . Like [38], reference [37] also defined the forward variable as a joint distribution of observation sequence and the underflow problem still existed as a result.

In this paper, to eliminate the underflow problem, the forward variable is instead modeled as the joint posterior probability of state at time t , corresponding sojourn time and the duration of last state on condition that the observation sequence $o_{1:t}$ is known:

$$\alpha_t(d^{'};i,d) \triangleq P\left\{q_{|t-d+1:t} = s_i, q_{|t-d-d^{'}+1:t-d|} = s_{i-1} \mid o_{1:t}, \lambda_d\right\} \tag{11}$$

where $q_{|t-d+1:t} = s_i$ means that the state becomes s_i at time $t-d+1$ and the state s_i has lasted for d at time t . Different from $q_{|t-d+1:t|} = s_i$, the $q_{|t-d+1:t} = s_i$ does not require that the state at time $t+1$ must change. As it is not necessary to infer whether the current state will change at the next time in the on-line estimation process, the form $q_{|t-d+1:t} = s_i$ is more convenient for the on-line state estimation.

Auxiliary forward variable at time t is defined as the joint posterior probability on condition that observation sequence $o_{1:t-1}$ is known:

$$J_t(d^{'};i,d) \triangleq P\left\{q_{|t-d+1:t} = s_i, q_{|t-d-d^{'}+1:t-d|} = s_{i-1} \mid o_{1:t-1}, \lambda_d\right\} \tag{12}$$

And there is:

$$\begin{aligned} & J_t(d^{'};i,d)c_i(o_t) \\ & = P\left\{q_{|t-d+1:t} = s_i, q_{|t-d-d^{'}+1:t-d|} = s_{i-1}, o_t \mid o_{1:t-1}, \lambda_d\right\} \end{aligned} \tag{13}$$

Then the one-step observation probability could be obtained as:

$$p\left\{o_t \mid o_{1:t-1}, \lambda_d\right\} = \sum\nolimits_{d^{'},i,d} J_t(d^{'};i,d)c_i(o_t) \tag{14}$$

Forward variable $\alpha_t(d^{'},i,d)$ could be seen as the normalization of $J_t(d^{'},i,d)c_i(o_t)$ with one-step observation probability as normalization factor:

$$\alpha_t(d^{'};i,d) = J_t(d^{'};i,d)c_i(o_t) \big/ p\left\{o_t \mid o_{1:t-1}, \lambda_d\right\} \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

As depicted in the HSMM for degradation process, the one-step transition is determined not only by the current state and corresponding lasted time but also the duration of preceded state. So the one-step transition probability could be defined as:

$$\begin{aligned}&\mu(d^{'},i,d;\,j)\\ &\triangleq P\left\{q_t = s_j \mid q_{[t-d:t-1} = s_i, q_{[t-d-d^{'}:t-d-1]} = s_{i-1}, \lambda_d\right\}\end{aligned} \tag{16}$$

When the state is $i>1$, for the progressive characteristic of degradation process, there is:

$$\mu(d^{'},i,d;\,j) = \begin{cases}\sum_{k\geq d+1} a_{d^{'}(i,k)} \Big/ \sum_{k\geq d} a_{d^{'}(i,k)} & j=i\\ a_{d^{'}(i,d)} \Big/ \sum_{k\geq d} a_{d^{'}(i,k)} & j=i+1\end{cases} \tag{17}$$

When the state is final degradation state $i=N$, according to (6) and (17), the one-step transition probability is $\mu(d^{'},N,d;N)=1$.

When the state is $i=1$, the duration of preceded state does not exist. To ensure coincident mathematical description, an arbitrary value x is used to denote the non-existent duration before initial degradation state. The one-step transition is:

$$\mu(x,i,d;\,j) = \begin{cases}\sum_{k\geq d+1} p_1(k) \Big/ \sum_{k\geq d} p_1(k) & j=1\\ p_1(d) \Big/ \sum_{k\geq d} p_1(k) & j=2\end{cases} \tag{18}$$

According to the definition of auxiliary forward variable and forward variable, the recursive formula from forward variable at time $t-1$ to auxiliary forward variable at time t could be obtained:

$$J_t(d^{'};i,d) = \begin{cases}\alpha_{t-1}(d^{'};i,d-1)\mu(d^{'},i,d-1;i) & d\neq 1\\ \sum_{d^{"}} \alpha_{t-1}(d^{"};i-1,d)\mu(d^{"},i-1,d^{'};i) & d=1\end{cases} \tag{19}$$

When the state at time t is initial state, the lasted time must be t and the duration of preceded state is an arbitrary value. In this situation, the forward variable is denoted by $\alpha_t(x;1,t)$ and auxiliary forward variable is denoted by $J_t(x;1,t)$. The initial condition of forward variable could be set as $\alpha_1(x;1,1)=1$.

Backward variable at time t is defined as the ratio between the joint posterior of state at time t, corresponding sojourn time and the duration of last state on condition that observation sequence $o_{1:T_l}$ is known and the forward variable at time t:

$$\beta_t(d^{'};i,d) \triangleq \eta_t(d^{'};i,d) \big/ \alpha_t(d^{'};i,d) \tag{20}$$

where:

$$\eta_t(d^{'};i,d) \triangleq P\left\{q_{[t-d+1:t} = s_i, q_{[t-d-d^{'}+1:t-d]} = s_{i-1} \mid o_{1:T_l}, \lambda_d\right\} \tag{21}$$

The backward recursion formula is:

$$\begin{aligned}\beta_t(d^{'};i,d) = &\left\{\mu(d^{'},i,d;i)c_i(o_{t+1})\beta_{t+1}(d^{'};i,d+1)\right.\\ &\left.+\,\mu(d^{'},i,d;i+1)c_{i+1}(o_{t+1})\beta_{t+1}(d;i+1,1)\right\}\\ &\times\left(p\left\{o_{t+1} \mid o_{1:t},\lambda_d\right\}\right)^{-1}\end{aligned} \tag{22}$$

The initial condition of backward recursion is:

$$\beta_{T_l}(d^{'};i,d) = 1 \tag{23}$$

The proof of backward recursion is given in appendix. The modified forward-backward algorithm is concluded in table 1.

Table 1: Modified forward-backward algorithm

The forward recursion:

1) For $t=1$, set initial forward variable $\alpha_1(x;1,1)=1$.

2) For $t>1$, calculate the auxiliary forward variable by (19), the one-step observation probability by (14) and forward variable by (15).

All the one-step observation probabilities need to be stored for backward recursion calculation.

The backward recursion:

1) For $t=T_l$, set initial backward variable $\beta_{T_l}(d^{'};i,d)=1$.

2) For $t<T_l$, calculate the backward variable by (22)

B. Degradation State Estimation

The two basic tasks of on-line health monitoring are the degradation state estimation and the remaining useful life estimation with the information inferred from the current and historical observations. The modified forward-backward algorithm is utilized for this purpose.

At current time T, the observation sequence is denoted by $o_{1:T}$. According to the definition of forward and backward variable, the posterior distribution of state at time $t\leq T$ could be written as:

$$\begin{aligned}P(q_t = s_i \mid o_{1:T}) &= \sum_{d^{'},d} \eta_t(d^{'};i,d)\\ &= \sum_{d^{'},d} \alpha_t(d^{'};i,d)\beta_t(d^{'};i,d)\end{aligned} \tag{24}$$

Because the initial condition of backward variable is $\beta_T(d^{'};i,d)=1$, the distribution of current state could be represented only by the forward variables at current time:

$$P(q_T = s_i \mid o_{1:T}) = \sum_{d,d^{'}} \alpha_T(d^{'};i,d) \tag{25}$$

Forward variables at current time could be calculated according to forward recursion listed in table 1. In fact, only one step forward recursion needs to be conducted when the forward variable at preceded observation time is saved. Hence, the on-line estimation process equals the on-line forward recursion process.

To predict state at future time $t>T$, the forward variables at time $t>T$ need to be calculated. Because the future observations have not been obtained at current time, the future observation could be set as an arbitrary value with uniform observation distribution. Then the auxiliary forward variable equals forward variable at future time and the forward recursion at future time could be simplified as:

$$\alpha_t(d^{'};i,d) = \begin{cases}\alpha_{t-1}(d^{'};i,d-1)\mu(d^{'},i,d-1;i) & d\neq 1\\ \sum_{d^{"}} \alpha_{t-1}(d^{"};i-1,d^{'})\mu(d^{"},i-1,d^{'};i) & d=1\end{cases} \tag{26}$$

With forward variables calculated at current time as initial condition, according to $t-T$ step forward recursion (26), the forward variables at future time $t>T$ could be obtained. Then the posterior distribution of state is obtained:

$$P(q_t = s_i \mid o_{1:T}) = \sum_{d,d^{'}} \alpha_t(d^{'};i,d) \quad t>T \tag{27}$$

Given the distribution of estimated state, diagnostic analysis could be conducted. In the applications, if the degradation state is divided uniformly and each state corresponds to a specific degradation value, an exact degradation state will be obtained and decided. In most cases, each degradation state corresponds to a range of degradation values, and under this condition the state with the maximum posterior probability is determined to be the final state.

C. Remaining Useful Life (RUL) Estimation

It has been pointed out [41] that current HSMM (HMM)-based prognosis methods could only estimate limited information, such as mean and variance of RUL, and cannot provide more statistics of the RUL. Recent extended works [32] [33] [37] did not solve this limitation however. In practice, it is important to compute the RUL distribution as it provides comprehensive information and confidence on the prognosis. In this section, the posterior distribution of RUL is derived via the proposed HSMM.

Remaining useful life is defined as the time interval between current time and the time at which the state reaches to final degradation state. According to the HSMM for degradation process, the distribution of RUL is determined by these three variables: current state i , corresponding lasted time d and the duration of last state $d^{'}$. Hence, these three variables are named as RUL factors in this paper. Remaining useful life is denoted by R and the duration of state k denoted by d_k . When the RUL factors are given, the RUL could be written as:

$$R=\sum_{k=i}^{N-1} d_k - d \tag{28}$$

The joint conditional distribution of variables $d_{N-1}\ldots d_{i+1}, d_i$ on condition that the RUL factors are given could be written as:

$$\begin{aligned} & P(d_i, d_{i+1}...d_{N-1} \mid d^{'}, i, d) \\ & = P(d_i \mid d_i \ge d, d_{i-1}=d^{'})P(d_{i+1} \mid d_i)...P(d_{N-1} \mid d_{N-2}) \\ & = \left(a_{d^{'}(i,d_i)} \Big/ \sum_{k\ge d} a_{d^{'}(i,k)} \right) \prod_{k=i}^{N-2} a_{d_k(k+1,d_{k+1})} \end{aligned} \tag{29}$$

Then the conditional distribution of RUL $P(R \mid d^{'}, i, d)$ could be obtained by the joint distribution in (29).

$$P(R \mid d^{'}, i, d) = \sum_{d_i,..d_{N-1}} P(d_i, d_{i+1}...d_{N-1} \mid d^{'}, i, d) \tag{30}$$

where $d_i + d_{i+1} + ... + d_{N-1} - d = R$.

The conditional distribution of RUL is only related to the parameters of HSMM, so it could be calculated and saved off-line after the HSMM is trained. The forward variables at current time could be seen as the posterior distribution of RUL factors at current time. With the conditional distribution of RUL calculated and saved off-line, the posterior distribution of RUL at current time is obtained:

$$P(R \mid o_{1:T}) = \sum_{d^{'},i,d} P(R \mid d^{'}, i, d)\alpha_T(d^{'}; i, d) \tag{31}$$

It is noted that the computation of the conditional distribution of RUL by (29) and (30) is very time-consuming and the numerical storage is huge especially when the maximum duration is large. While in many applications, only some posterior statistics rather than the posterior distribution of RUL need to be considered. The posterior statistics of RUL $T(R \mid o_{1:T})$ could be regarded as the weighted average of conditional statistics of RUL $T(R \mid d^{'}, i, d)$ with $\alpha_T(d^{'}; i, d)$ as weighted coefficient:

$$\begin{aligned} & T(R \mid o_{1:T}) \\ & = \sum_R T(R)\left\{ \sum_{d^{'},i,d} P(R \mid d^{'}, i, d)\alpha_T(d^{'}; i, d) \right\} \\ & = \sum_{d^{'},i,d} \left\{ \sum_R T(R) \sum_{d^{'},i,d} P(R \mid d^{'}, i, d) \right\} \alpha_T(d^{'}; i, d) \\ & = \sum_{d^{'},i,d} T(R \mid d^{'}, i, d)\alpha_T(d^{'}; i, d) \end{aligned} \tag{32}$$

Hence, it just needs to calculate and save some conditional statistics (such as conditional expectation and variance) rather than the conditional distribution of RUL off-line. This will reduce the storage space greatly. The computation cost could also be reduced by computing the conditional statistics via Monte Carlo simulation method [42].

IV. Experimental Validation

A. Case Study

Case study is conducted on tool wear monitoring in a high-speed CNC milling center. The flank wear of carbide ball-nose cutters with three flutes and cutting force signal are recorded in experiments by a Kistler dynamometer mounted under the workpiece. The Inconel 718 is used as work-piece material, which is a typical hard-to-machine aerospace material. The spindle rotation frequency is 172 Hz and sampling rate is 10 kHz. There are in total 23 tests carried out, with varying working parameters in feed rate and depth of cut. Every experiment lasts for 105 passes and the time interval of each cutting pass is 12s.

Tool wear is defined as the mean flank wear of three flutes. Tool wear state is divided into 5 states according to the range of tool wear (shown in table 2). The tool wear is measured during the recess of tool holder. Thirteen tests are used as training samples and the remaining 10 tests are used as testing samples. In both of the training and testing tests, tool wear is measured and each cutting pass is labeled the wear state according to Table 2.

Table 2: Tool wear state

State label	1	2	3	4	5
Wear range(μm)	0-60	60-90	90-120	120-150	>150

The number of cutting passes is utilized to represent the length of time for duration and RUL. For example, the duration of state 1 is 5 means state 1 will lasts for 5 cutting passes. Likewise, the RUL of a tool is 20 means the tool will reaches to the final state 5 after 20 cutting passes. As the sequence of tool wear state of testing test is recoded in experiments, the durations and the cutting pass at which the tool reaches state 5 could conveniently be determined. The real RUL at certain cutting pass is the number of cutting passes between the cutting pass and the cutting pass at which the tool becomes state 5. The sequence of real RUL of testing test is also recorded and then used as a reference of RUL estimation in testing process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

B. Feature Extraction and Quantization

Related researches show that the distribution of energy of cutting force signal in different frequency bands is capable to reflect the wear state. Cutting force signal could be approximately seen as a periodic signal with the spindle rotation frequency as fundamental characteristic frequency (172 Hz).

Wavelet packet decomposition (WPD) [43] is utilized to extract the features in different frequency bands of cutting force signal. In order to separate information inferred by different harmonic components of cutting force signal into different elements of feature vector, the frequency band should be divided into at least 59 segments (10 kHz/ 172 Hz). To meet this need, cutting force signal is decomposed into 64 frequency bands via 6 level db8 wavelet package decomposition of cutting force signal. The mean energy of one frequency band is defined as the ratio between the sum of square of wavelet packet coefficients $d_{j,k}^n$ and the number of wavelet packet coefficients in one frequency band:

$$E_j = \frac{1}{N_j}\sum_{k=1}^{N_j}\left|d_{j,k}^n\right|^2 \tag{33}$$

where $d_{j,k}^n$ is obtained by the wavelet packet transform of the signal localized at $2^j k$ in scale 2^j, and n is the oscillation parameter $n = 1, 2...$. Then the feature vector consists of mean energy in 64 frequency bands. Cutting force signal extracted from one cutting pass and its mean energy in each wavelet packet are shown in Figure 2.

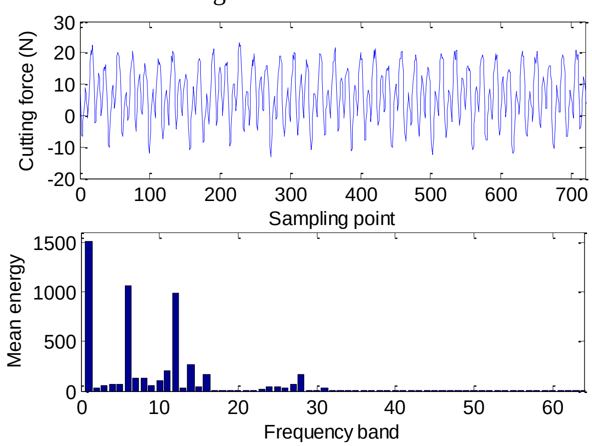


Fig.2. Cutting force signal and its energy in wavelet packet

As introduced in section II, there exists two ways to build the generative model for high-dimension observation feature. One way is to assign an explicit distribution for the features, and the other way is to quantize the features to obtain discrete observations. For the high dimensional feature vectors in this study, it is difficult to assign an explicit distribution form for them. So the quantized approach is adopted. The discrete observation is used as the observation of HSMM and the probability distribution of discrete observation is used to descript the relationship between tool wear state and observation feature.

To obtain discrete observation, k-means approach [44] is adopted to quantize the wavelet packet feature vector. The total 1365 wavelet packet feature vectors extracted from 13 training tests are divided into 20 clusters with k-means method. The quantized result (discrete observation) is the cluster to which the wavelet packet feature vectors belongs to. For testing samples, the Euclidean distance from the wavelet packet features to each cluster's center are computed, and then the discrete observation is assigned to the cluster with the minimum distance.

Figure 3 shows the discrete observations of training test 1 and testing test 1. With all of the discrete observations and states in training set, the probability distribution of discrete observation in each state could be estimated, which are further studied in the next sections.

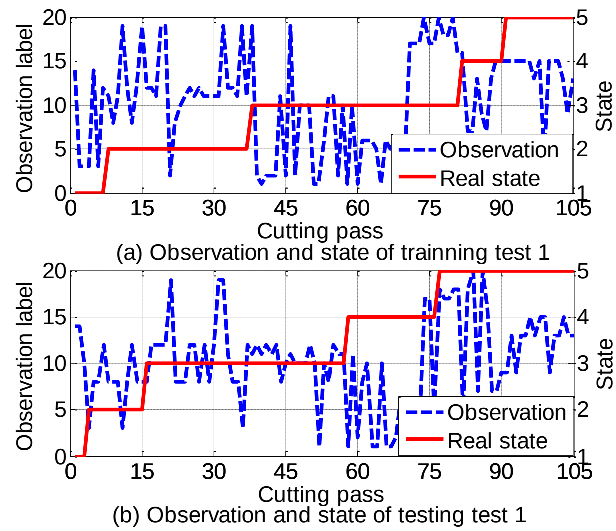


Fig.3 Evolution of discrete observation and tool wear state.

C. Training of HSMM for Tool Wear Monitoring

Apparently, the estimated probability $\hat{c}_i(v_j)$ is the frequency that discrete observation $j \in \{1, 2, ..., 20\}$ occurs when the wear state is $i \in \{1, 2, ..., 5\}$ in the training experiments. The estimated observation probability is given in Figure 4.

The distribution of discrete observation could be used to estimate the tool wear state. For example, given observation 3, the state will be estimated as state 1 as the probability that observation 3 belongs to state 1 is bigger than the others (as Figure 4 shows). Because the method does not utilize any prior knowledge of durations, it is named as distribution method without duration in this paper. Obviously, the method without duration is a kind of pure data-based method. The diagnosis result via the method without duration will be given in section IV-D.

The duration and tool life are largely determined by cutting conditions. For different cutting conditions, the duration may be quite different. Hence, there must exist a high modeling error if a fixed HSMM is trained for different cutting conditions. To reduce the model uncertainty, the Taylor's tool life [45] is referred for the tool wear modeling. The two-dimensional normal distribution is utilized to represent the dependence of adjacent durations.

$$\begin{bmatrix} d_i \\ d_{i+1} \end{bmatrix} \sim N\left(T_e\begin{bmatrix} \mu_i \\ \mu_{i+1} \end{bmatrix}, T_e^2\begin{bmatrix} \sigma_i^2 & \tau_{i,i+1} \\ \tau_{i,i+1} & \sigma_{i+1}^2 \end{bmatrix}\right) \tag{34}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

where T_e represents the Taylor tool life. The Taylor tool life of training test p is denoted by T_e^p and the duration of wear state i in training test p is denoted by d_i^p .The total number of training tests is $K = 13$. The parameters in (34) are estimated as:

$$\hat{\mu}_i = \frac{1}{K}\sum_{p=1}^{K}\frac{d_i^p}{T_e^p} \tag{35}$$

$$\hat{\sigma}_i^2 = \frac{1}{K}\sum_{p=1}^{K}\left(\frac{d_i^p}{T_e^p} - \hat{\mu}_i\right)^2 \tag{36}$$

$$\hat{\tau}_{i,i+1} = \frac{1}{K}\sum_{p=1}^{K}\left(\frac{d_{i+1}^p}{T_e^p} - \hat{\mu}_{i+1}\right)\left(\frac{d_i^p}{T_e^p} - \hat{\mu}_i\right) \tag{37}$$

The discrete transition probability in (5) and initial probability in (7) could be obtained from (34). If the covariance $\tau_{i,i+1}$ in (37) is set as zero, the HSMM with dependent durations becomes the HSMM with independent durations. As a result, the HSMM with independent durations could be regarded as a special HSMM with dependent durations. Hence, the proposed on-line monitoring approach could also be applied when the HSMM with independent durations is utilized to model the tool wear process.

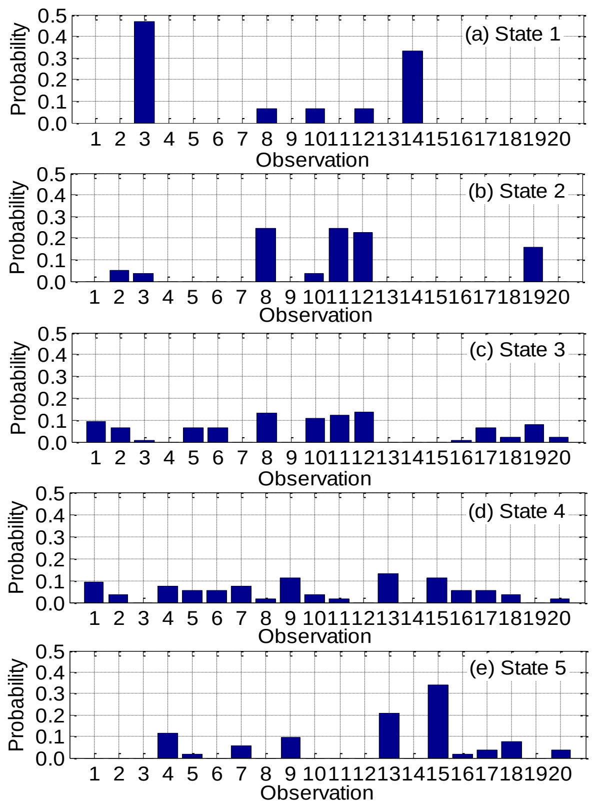


Fig.4. Estimated discrete observation distribution

The estimated parameters have great influence on the diagnosis and prognosis results. If the expectation $\hat{\mu}_i$ is bigger than the real expectation, for most of the testing tests, the evolution of estimated degradation state will lag behind the evolution of real state and the estimated RUL will be larger than the real RUL. This will cause the problem of missing alarm. On the contrary, if the expectation is smaller than the real expectation, the estimated RUL will be smaller than the real one and the false alarm will occur. The covariance $\hat{\tau}_{i,i+1}$ represents the correlation between the durations. Positive (negative) covariance means positive (negative) correlation between durations. If the estimated covariance is contrary to the real covariance, the estimated RUL will seriously deviate from the real RUL. Variance $\hat{\sigma}_i^2$ reflects the confidence level of duration. Larger variance corresponds to smaller confidence. If the variance is set high, the information inferred from durations will have less influence on the diagnosis results and the diagnosis results will be determined by the observations more. If the variance is set small, the monitoring result will be more determined by the duration information and the estimated RUL will approximates to the prior RUL expectation.

As the HSMM is trained from training set and the modeling accuracy is largely determined by the number of training samples, abundant training samples will be necessary and important for the accurate modeling. In this study, the HSMM is trained from the 13 training tests, and it is sufficient to represent the 10 testing tests. This will further be verified by the monitoring results presented in the next section.

D. Diagnosis and Prognosis Results

1) Diagnosis and Prognosis via HSMM Approach

In this section, the diagnosis and prognosis are shown for the on-line tool wear monitoring process via HSMM method. At each cutting pass, based on the trained HSMM in section above, the tool wear state and RUL are inferred from the discrete observations. The evolution of diagnostic wear state is shown in Figure 5. Figure 5 shows that utilizing of duration information leads to more accurate diagnostic results and the HSMM with dependent durations has more accurate diagnostic results than HSMM with independent durations. This could also be verified by the recognition rate of all tests listed in Table 3.

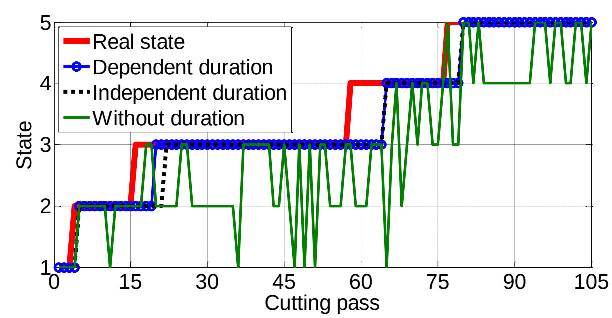


Fig.5. Estimated wear state

Under the premise of correct estimation, higher confidence level shown as a more centralized distribution gives decision-maker more confident diagnostic results. The evolution of the posterior distribution of state is shown in Figure 6. From the figure, it could be found that the distribution of estimated state is more centralized with dependent duration than that with independent durations when the wear state is estimated correctly (around pass 55). So it could be concluded that the fuller use of dependent characteristic of adjacent durations leads to more confident diagnosis.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

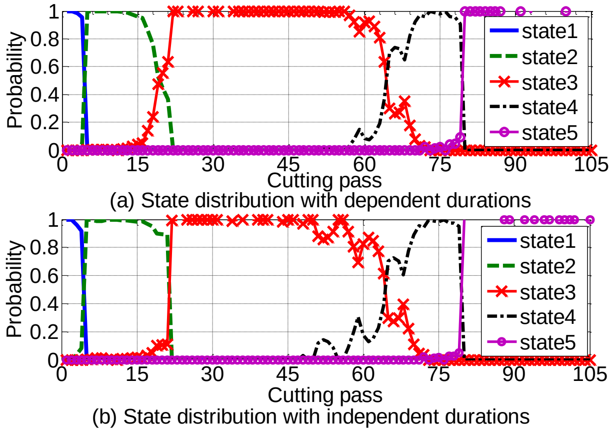


Fig.6. Evolution of posterior probability distribution of wear state

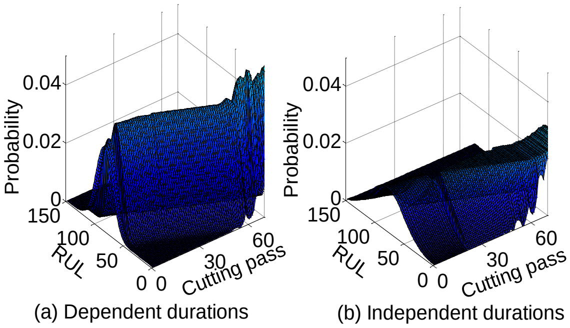


Fig7. Evolution of distribution of RUL

Figure 7 gives the evolution of distribution of RUL at the passes in which the tool wear state is not estimated as final state (when the estimated state is final state, the probability that RUL is zero approximates to one). It shows that estimated RUL has a more centralized distribution with dependent duration than that with independent duration. Namely, it leads to a more confident RUL estimation with dependent durations. Estimated error of RUL which reflects the accuracy of RUL estimation is defined as the root of mean squared error. The evolutions of error and mean of estimated RUL are given in Figure 8. It indicates that the RUL estimation with dependent durations is more accurate as the estimation error with dependent durations is less than that with independent durations. Figure 9 shows the evolution of optimal confidence interval at the 90% confidence level. It also shows that the RUL estimation with dependent durations is more confident as the width of confidence interval with dependent durations is narrower than that with independent durations. Therefore, we could say that RUL estimation becomes more accurate and more confident when the dependent relationship of durations of adjacent wear state is included in the HSMM for tool wear process.

Figure 8 shows that the estimated RUL (mean RUL) is higher than the real RUL and missing alarms may occur. To eliminate this problem, a lower confidence bound of RUL is set as the estimated RUL after cutting pass 65 at which the estimated state becomes last-to-second state (state 4). As the lower bound of RUL is smaller than the real RUL (Figure 9 shows), the RUL of the tool will be estimated as zero before the tool actually reaches to the final wear state. In this way, the final tool wear state could be avoided and the process is relieved from loss due to worn tool.

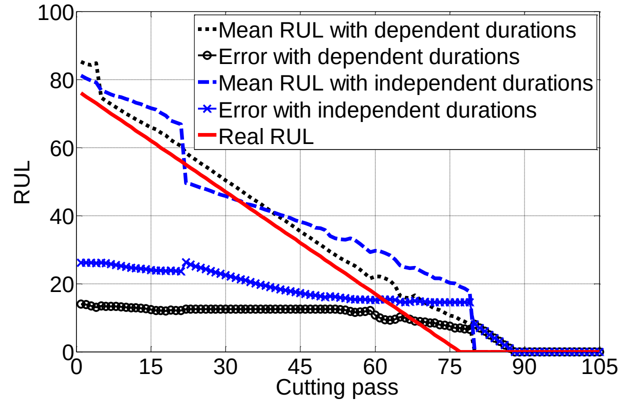


Fig.8.The statistics of RUL

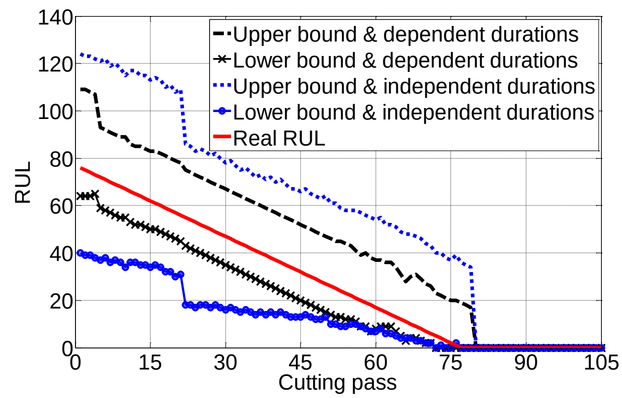


Fig.9. Optimal confidence interval at 90% confidence level

2) Comparison to Pure Data-based Approach

As classical pure-data monitoring methods, SVM and backpropagation neural network (BPNN) are applied to infer the tool wear state. The extracted wavelet packet features of cutting force are used as the input of SVM and BPNN. The training and testing sets used in SVM and BPNN approaches are the same as the sets used in the proposed HSMM approach. Generally, SVM is applied for binary classifications. To recognize 5 tool wear states, ten SVMs are trained, with each SVM separates two tool wear states. In testing process, the wavelet feature is applied to all of the 10 SVMs and each SVM outputs a state. Then the estimated state is set as the state which has maximum occurrence in the outputs of the 10 SVMs. With empirical studies, the trained BPNN has three hidden layers and the number of nodes at each layer is 32, 16 and 8.

As 10 tests are used as testing samples and each test has 105 cutting passes, the total number of testing samples is 1050. The total recognition rate is defined as the ratio between the number of state estimated correctly and the total number 1050. Table 3 lists the total recognition rates of all testing tests as well as the recognition rate for each of the state. The RUL error rate is defined as the ratio between RUL error and the real RUL. Table 3 also lists the average RUL error rate (ARER) of the testing samples. It also implies that the fuller use of the latent evolution law of hidden state leads to more accurate prognosis results.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

Table 3: Recognition rate and ARER of all tests

Approach	State 1	State 2	State 3	State 4	State 5	Total	ARER
Dependent duration	0.8810	0.8773	0.9278	0.9270	0.9066	0.9143	0.1269
Independent duration	0.7619	0.7975	0.8763	0.7584	0.8132	0.8286	0.2986
Without duration	0.6429	0.6626	0.5711	0.5169	0.5659	0.5781	NA
SVM	0.5476	0.5890	0.8247	0.6067	0.5879	0.6990	NA
BPNN	0.4762	0.6196	0.8866	0.6348	0.6429	0.7431	NA

Table 3 shows that the diagnosis result via data model-integrated method is more accurate than those via pure-data method. In pure data-based approaches (without duration, SVM and BPNN), only the wavelet packet feature is utilized to estimate the tool wear. As different tool state may produce similar features and the feature space of different tool state overlaps, the recognition error would be high when overlapped wavelet packet feature occurs. HSMM approaches not only model the relationship between hidden states and features but also include the evolution model of states. In this study, both of the observation and the evolution law of hidden state are taken into account. This will reduce the risk of misclassification. For example, at cutting pass 90 of testing test 1, observation 9 occurs (shown in Figure 3b), the distribution without duration utilizes the relationship between the observation and state (Figure 4) to estimate the state and the state is wrongly estimated as state 4. The same case occurs when the SVM or BPNN method is adopted. With the fact that the previous state is most likely state 5 and the evolution law that state 5 will not change, the state is correctly estimated as state 5 when the proposed HSMM method is adopted.

Table 3 shows that the recognition rate of state 3 in SVM and BPNN methods is much higher than other states. This is because there is much more training samples of state 3 in the training set. On the contrary, the other four states could not be efficiently detected by the trained SVM and BPNN as there are much less training samples of these four states. While the generative model is used as the observation model and the duration of state is included, the HSMM method is less influenced by the number of training samples in different states. This could be verified by the fact that the recognition rate for different states does not vary much when the HSMM method is adopted.

3) On the Adaptability of the Proposed HSMM Approach

The experimental results imply that the proposed HSMM approach is applicable in progressive tool wear monitoring and tool life estimation. As the HSMM is built from historical data set, abundant historical data set is necessary for the accurate diagnosis and prognosis. For other applications with progressive degradation process, the proposed HSMM approach is applicable provided with enough training data. It is noted that the degradation process is assumed to change progressively in the model and the proposed HSMM is not applicable in monitoring degradation process with abrupt changes such as chipping of tool and the fracture of bearing.

V. Conclusions

In this study, an improved hidden semi-Markov process is proposed to model the degradation process with dependent durations. A modified forward-backward algorithm which is able to avoiding underflow problem is also developed. And based on the modified forward-backward algorithm, on-line degradation state estimation and remaining useful prediction could be conducted conveniently. Case studies have shown that by including the dependence of durations in the HSMM it leads to more accurate and more confident estimation results, with the increase of both the state recognition rate and RUL estimation. And notably the confidence of RUL has a great improvement. In further work, the jumping of HSMM states should be investigated in HSMM so that the abrupt changes of equipment conditions could also be modeled and predicted.

APPENDIX: Proof of the Backward Recursion Formula

For mathematical convenience, we use notation $e_t(i,d,d^{'})$ to represent the events that state at time t is s_i, corresponding lasted time is d and duration of preceded state is $d^{'}$.

$$e_t(i,d,d^{'}) \triangleq \left\{q_{|t-d+1:t} = s_i, q_{|t-d-d^{'}+1:t-d|} = s_{i-1}\right\} \tag{38}$$

According to the definition of backward variable, backward variable could be rewritten as:

$$\begin{aligned}\beta_t(d^{'};i,d) &= \eta_t(d^{'};i,d) \big/ \alpha_t(d^{'};i,d)\\ &= \frac{P\left\{e_t(i,d,d^{'}), o_{1:t}, o_{t+1:T_l}\right\} P(o_{1:t})}{P\left\{e_t(i,d,d^{'}), o_{1:t}\right\} P\left(o_{1:T_l}\right)}\\ &= P\left\{o_{t+1:T_l} \mid e_t(i,d,d^{'}), o_{1:t}\right\}\left\{P(o_{1:t}) \big/ P(o_{1:T_l})\right\}\end{aligned} \tag{39}$$

When the current state, corresponding lasted time and duration of preceded state are given, the future observations are independent to the current and historical observations. Therefore, equation (39) could be written as:

$$\beta_t(d^{'};i,d) = P\left\{o_{t+1:T_l} \mid e_t(i,d,d^{'})\right\}\left\{P(o_{1:t}) \big/ P(o_{1:T_l})\right\} \tag{40}$$

The probability of future observations could be decomposed into two parts according to whether the state changes or not:

$$\begin{aligned}&P\left\{o_{t+1:T_l} \mid e_t(i,d,d^{'})\right\}\\ &= \mu(d^{'};i,d;i) c_i(o_{t+1}) P\left\{o_{t+2:T_l} \mid e_{t+1}(i,d+1,d^{'})\right\}\\ &+ \mu(d^{'};i,d;i+1) c_{i+1}(o_{t+1}) P\left\{o_{t+2:T_l} \mid e_{t+1}(i+1,1,d)\right\}\end{aligned} \tag{41}$$

Multiplying (40) by $P(o_{1:t+1}) \big/ P(o_{1:T_l})$ and then taking (39) into the equation, we get the backward recursion form:

$$\begin{aligned}&\beta_t(d^{'};i,d)\{P(o_{t+1}\mid o_{1:t})\}\\&=\mu(d^{'};i,d;i)c_i(o_{t+1})\beta_{t+1}(d^{'};i,d+1)\\&+\mu(d^{'};i,d;i+1)c_{j+1}(o_{t+1})\beta_{t+1}(d;i+1,1)\end{aligned} \tag{42}$$

ACKNOWLEDGEMENT

This work was supported in part by the CAS 100 Talents Program, Chinese Academy of Sciences, and in part by the National Natural Science Foundation of China under Grant 51475443.

We would also like to thank the anonymous reviewers on their constructive comments and suggestions on improving the quality of this paper.

REFERENCES

[1] R. Teti, K. Jemielniak , G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Ann-Manuf Technol.*, vol. 59, no. 2, pp. 717–739. 2010.

[2] D. F. Shi, N. N. Gindy, "Industrial applications of online machining process monitoring system," *IEEE-ASME Trans. Mechatron.*, vol. 12, no. 5, pp. 561–564, Oct. 2007

[3] R. Isermann. "Model-based fault-detection and diagnosis–status and applications," *Annu. Rev. Control*, vol.29, no.1, pp.71-85. Dec. 2005.

[4] S. Yin, X. W. Li, H. J. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: an overview," *IEEE Trans .Ind. Electron.,* vol. 62, no.1, pp.657-667. Jan. 2015.

[5] X. S. Si, W. B. Wang, C. H. Hu, M. Y. Chen, and D. H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mech. Syst. Signal Process*, vol. 35, no.1- 2, pp. 219–237, Feb. 2013.

[6] M. Braglia, and D. Castellano, "Diffusion theory applied to tool-life stochastic modeling under a progressive wear process," *J. Manuf. Sci. Eng.-Trans. ASME*, vol.136, no.3, pp.1-12, Mar. 2014.

[7] H. C. Yan, J. H. Zhou, and C. K. Pang, "Gamma process with recursive MLE for wear PDF prediction in precognitive maintenance under aperiodic monitoring," *Mechatronics*, vol. 31, pp. 68–77, Oct. 2015.

[8] N. P. Li, Y. G. Lei, J, Lin, and S. X. Ding, "An improved exponential model for prediction remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.

[9] W. Wang, and O. A. Jianu, "A smart sensing unit for vibration measurement and monitoring," *IEEE-ASME Trans. Mechatron.*, vol. 15, no.1, pp. 70–78, Feb. 2010.

[10] K. P. Zhu, T. Mei, and D. S. Ye, "Online condition monitoring in micromilling: A force waveform shape analysis approach," *IEEE Trans .Ind. Electron.*, vol. 62, no. 6, pp. 3806–2813, Jun. 2015.

[11] M. Van, and H. J. Kang. "Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization". *IEEE Trans. Ind. Informat*, vol.12, no.1, pp. 124–135, Feb. 2016.

[12] Z. Du, X. Chen, H. Zhang, and R. Q. Yan, "Sparse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6594–6605, Oct. 2015.

[13] J. H. Zhou, C. K. Pang, Z. W. Zhong, and F. L. Lewis, "Tool wear monitoring using acoustic emissions for dominant-feature identification," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 547–559, Jan. 2011.

[14] A. Widodo, B. S. Yang. "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process*, vol. 21, no.6, pp. 2560-2574, Jan. 2007.

[15] S. N. Huang, K. K. Tan, and M. B. Xiao, "Automated fault diagnosis and accommodation control for mechanical systems," *IEEE-ASME Trans. Mechatron.*, vol. 20, no. 1, pp. 155–165, Feb. 2015.

[16] A. P. Ompusunggu, J. M. Papy, and S. Vandenplas, "Kalman-filtering-based prognostics for automatic transmission clutches," *IEEE-ASME Trans. Mechatron.*, vol.21, no.1, pp. 419 – 430, Feb. 2016.

[17] Y. Wang, Y. Z. Peng, Y. Y. Zi, X. H. Jin, and K. L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informat.*, vol. 12, no.3, pp. 924–932, Jun. 2016.

[18] A. Soualhi, G. Clerc, H. Razik, M. E. Badaoui, and F. Guillet "Hidden Markov models for the prediction of impending faults," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 1781–1790, May. 2016.

[19] E. Athanasopoulou, L. Li, and C. N. Hadjicostis, "Maximum likelihood failure diagnosis in finite state machines under unreliable observations," *IEEE Trans. Autom. Control*, vol.55, no.3, pp. 579-593, Mar. 2010.

[20] S. Lee, L. Li, and J. Ni, "Online degradation assessment and adaptive fault detection using modified hidden Markov model," *J. Manuf. Sci. Eng.-Trans. ASME* , vol. 132, no.2, pp.1-11, Apr. 2010.

[21] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," *Int. J. Prod. Res.*, vol. 43, no. 6, pp. 1275–1293, Mar. 2005.

[22] R. K. Fish, M. Ostendorf, G. D. Bernard, and D. A. Castanon, "Multilevel classification of milling tool wear with confidence estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 75–85, Jan. 2003.

[23] L. T. Wang, M. G. Mehrabi, and E. Jr, "Hidden Markov model-based tool wear monitoring in turning," *J. Manuf. Sci. Eng.-Trans. ASME*, vol. 124, no. 3, pp. 651–658, Aug. 2002.

[24] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering," *IEEE Trans. Ind. Electron.*, vol.58, no.9, pp.4353-4364, Sep. 2011.

[25] J. L. Zhu, Z. Ge, and Z. Song, "HMM-driven robust probabilistic principal component analyzer for dynamic process fault classification," *IEEE Trans. Ind. Electron.*, vol.62, no.6, pp.3814-3821. Jun. 2015.

[26] O. Geramifard, J. X. Xu, J. H. Zhou, and X. Li, "A physically segmented hidden Markov model approach for continuous tool condition monitoring: Diagnostics and prognostics," *IEEE Trans. Ind. Informat*., vol. 8, no. 4, pp. 964–973, Nov. 2012.

[27] O. Geramifard, J. X. Xu, J. H. Zhou, and X. Li, "Multimodal hidden Markov model-based approach for tool wear monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2900–2911, Jun. 2014.

[28] H. Nakamura, M. Chihara, T. Inoki, T. Higaki, K. Okuda, and Y. Mizuno, "Impulse testing for detection of insulation failure of motor winding and diagnosis based on Hidden Markov Model," *IEEE Trns. Dielectr. Electr. Insul*, vol. 17, no.5, pp.1619-1627. Oct. 2010.

[29] J. Ying, T. Kirubarajan, K. R. Pattipati, and A. Patterson-Hine, "A hidden Markov model-based algorithm for fault diagnosis with partial and imperfect tests," *IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev*., vol.30, no.4, pp. 463-473. Nov. 2002.

[30] F. Camci and R. B. Chinnam, "Health-state estimation and prognostics in machining processes," *IEEE Trans. Autom. Sci. Eng*, vol. 7, no. 3, pp. 581–591, Jul. 2010.

[31] K. P. Zhu, Y. S. Wang, and G. S. Hong, "Multi-category micro-milling tool wear monitoring with continuous hidden Markov models," *Mech. Syst. Signal Process*, vol. 23, no. 2, pp. 547–560, Feb. 2009.

[32] Q. M. Liu, M. Dong, W. Y. Lv, X. L. Geng, and Y. P. Li, "A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis," *Mech. Syst. Signal Process.*, vol. 64–65, pp. 217–232, Dec. 2015.

[33] D. A. Tobon-Mejia, K. Maderjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models," *IEEE Trans. Reliab.*, vol.61, no.2, pp.491–503, Jun. 2012.

[34] O. Geramifard, J. X. Xu, J. H. Zhou, and X. Li, "Continuous health condition monitoring: A single hidden semi-Markov model approach," *in Proc. IEEE Int. Conf. Prognostics and Health Management (PHM)*, 2011, pp.1–10.

[35] X. Wu, Y. Li, and A. K. Guru. "Integrated prognosis of AC servo motor driven linear actuator using hidden semi-Markov models," *in Proc. IEEE Int. Conf. Electric Machines and Drives*, 2009, pp.1408-1413.

[36] Y. Peng, and M. Dong, "A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction," *Mech. Syst. Signal Process.*, vol.25, no.1, pp. 237-252. Apr. 2011.

[37] N. Wang, S.D. Sun, Z. Q. Cai, S. Zhang, and C. Saygin, "A hidden semi-Markov model with duration-dependent state transition probabilities for prognostics," *Math. Probl. Eng.*, pp.1-10, Apr. 2014.

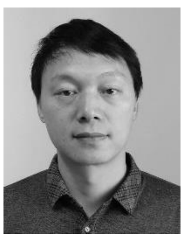
[38] S. Z. Yu, "Hidden semi-Markov models," *Artif. Intell*., vol.174, no.2, pp.215–243, Feb. 2010.

[39] A. Gersho and R. M. Gray, Vector quantization and signal compression, Springer Science & Business Media, 2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMECH.2018.2823320, IEEE/ASME Transactions on Mechatronics

[40] S. Z. Yu, and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden semi-Markov model," *IEEE Trans. Signal Process.*, vol.54, no.5, pp.1947–1951, May. 2006.
[41] X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," Eur. J. Oper. Res., vol. 213, no. 1, pp. 1–14, Aug. 2011.
[42] A. Doucet, N. D. Freitas, and N. Gordon. "An introduction to sequential Monte Carlo methods." Sequential Monte Carlo methods in practice. Springer New York, 2001.
[43] S.G. Mallat, A Wavelet Tour of Signal Processing, San Diego: Academic, 1998.
[44] A. K. Jain. "Data clustering: 50 years beyond K-means", *Pattern Recognit. Lett*, vol.31, no.8, pp. 651-666, 2010.
[45] J. Karandikar, A. Abbas, and T. Schmitz, "Tool life prediction using Bayesian updating, Part 1: Milling tool life model using a discrete grid method," Precis. Eng., vol. 38, no. 1, pp. 9–17, Jan. 2014.

Tongshun Liu is currently working toward the Ph.D degree in the Department of Automation, University of Science and Technology of China, Hefei, China. His research interests are in the areas of manufacturing automation, condition monitoring and statistical signal processing.



Kunpeng Zhu received the Ph.D. degree from the Department of Mechanical Engineering, National University of Singapore (NUS), Singapore, in 2007. He is a Professor and the Head of the Precision Manufacturing Laboratory at the Institute of Advanced Manufacturing Technology, Chinese Academy of Sciences, Changzhou, China. His research interests are in the areas of manufacturing automation, process modeling and monitoring, additive manufacturing and informatics.

Liangcai Zeng received the Ph.D. degree from the Department of Mechanical Engineering, Wuhan University of Science and Technology (WUST), China, in 2005. He is a Professor and the Dean of School of Machinery and Automation, Wuhan University of Science and Technology. His interests are in the fluid machinery process monitoring and fault diagnosis, electro-Hydraulic servo system and intelligent control.