A generalized interval probability-based optimization method for training generalized hidden Markov model

Fengyun Xie [a,b], Bo Wu [a], Youmin Hu [a,*], Yan Wang [c], Guangfei Jia [a], Yao Cheng [a]

[a] State Key Laboratory for Digital Manufacturing Equipment and Technology, Huazhong University of Science& Technology, Wuhan 430074, PR China
[b] School of Mechanical and Electronical Engineering, East China Jiaotong University, Nanchang 330013, PR China
[c] Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, USA

ARTICLE INFO

ABSTRACT

Recently a generalized hidden Markov model (GHMM) was proposed for solving the information fusion problems under aleatory and epistemic uncertainties in engineering application. In GHMM, aleatory uncertainty is captured by the probability measure whereas epistemic uncertainty is modeled by generalized interval. In this paper, the problem of how to train the GHMM with a small amount of observation data is studied. An optimization method as a generalization of the Baum–Welch algorithm is proposed. With a generalized Baum–Welch's auxiliary function and the Jensen inequality based on generalized interval, the GHMM parameters are estimated and updated by the lower and upper bounds of observation sequences. A set of training and re-estimation formulas are developed. With a multiple observation expectation maximization (EM) algorithm, the training method guarantees the local maxima of the lower and the upper bounds. Two case studies of recognizing the tool wear and cutting states in manufacturing is described to demonstrate the proposed method. The results show that the optimized GHMM has a good recognition performance.

1. Introduction

The hidden Markov model (HMM) with the capability of statistical learning and classification has been widely applied in speech recognition [1,2], character recognition [3] and fault diagnosis [4]. Yet the HMM does not differentiate two types of uncertainties. *Aleatory* uncertainty is inherent randomness and irreducible variability in nature, whereas *epistemic* uncertainty is reducible because it comes from the lack of knowledge. The sources of epistemic uncertainty cannot be ignored in engineering applications. All models have errors because approximations are always involved in model construction, and all experimental measurements contain systematic errors. In order to improve the robustness of analysis, the effect of epistemic uncertainty should be considered separately from the one from aleatory uncertainty. Given the very different sources of the two uncertainty components, we use two different forms to distinguish the two. Aleatory uncertainty is represented as probability, whereas interval is used to capture epistemic uncertainty. Intervals naturally capture the measurement errors, as well as the lower and upper bounds of model errors from the incomplete knowledge, without the assumptions of probability distributions.

Recently a generalized interval probability which combines generalized intervals with probability measures was proposed by Wang [5]. The generalized interval is used to represent the epistemic uncertainty component. Compared to the classical interval, generalized interval based on the Kaucher arithmetic [6] has better algebraic properties so

* Corresponding author. Tel./fax: +86 27 87541940.
E-mail address: youmhwh@mail.hust.edu.cn (Y. Hu).

that the calculus can be simplified. In addition, a generalized hidden Markov model (GHMM), as a generalization of HMM, was proposed for statistical learning and classification with both uncertainty components [7]. In GHMM, the precise values of a probability for HMM are replaced by the generalized interval probabilities.

Similar to HMM, the optimization of GHMM parameters is also the central problem in model calibration [8]. In this paper, an optimization method, which is based on a generalized Jensen inequality and a generalized Baum–Welch algorithm (GBWA) in the context of generalized interval probability theory, is proposed for training GHMM. The parameters of GHMM are estimated and updated by using GBWA. Different from the multiple observation training in HMM [9], the GHMM parameters are estimated and updated by the given lower and upper bounds of observation sequences. The lower and upper bounds capture the epistemic uncertainty associated with the observation, such as systematic error and bias. Based on a generalized Baum–Welch's auxiliary function, a set of training equations are developed by optimizing the objective function. A set of GHMM re-estimated formulas has been deduced by the unique maximum of the objective function. The proposed GBWA optimization method takes advantage of the good algebraic property in the generalized interval probability, which provides an efficient approach to train the GHMM. In order to demonstrate the performance of the proposed optimization method for training GHMM, two cases of tool state and cutting state recognition in manufacturing processes is provided. The tool states and cutting states are recognized by the GBWA training algorithm of GHMM.

In the remainder of this paper, Section 2 provides the overview of relevant work in generalized interval, generalized interval probability, and GHMM. Section 3 introduces a generalized Jensen inequality. Section 4 introduces optimization methods in training process of the GHMM. Section 5 demonstrates the application for the tool state and cutting state recognition based on the GBWA. Finally, Section 6 is the conclusion.

2. Background

2.1. Generalized interval

The generalized interval is an extension of the classical interval with better algebraic and semantic properties based on the Kaucher arithmetic. A generalized interval $\mathbf{x} := [\underline{x}, \overline{x}]$, $(\underline{x}, \overline{x} \in \mathbb{R})$ is defined by a pair of real numbers as $\underline{x}$ and $\overline{x}$ [10,11]. The generalized interval is not constrained by that the lower bound should be less than or equal to the upper bound. For instance, both [0.1, 0.3] and [0.3, 0.1] are valid in generalized interval. Interval [0.1, 0.3] is called *proper*, whereas interval [0.3, 0.1] is called *improper*. The relationship between proper and improper intervals is established with the operator *dual*, defined as $dual\,\mathbf{x} := [\overline{x}, \underline{x}]$. Operator *pro* returns the classical proper interval. For instance, $pro[0.3, 0.1] = [0.1, 0.3]$, and $pro[0.1, 0.3] = [0.1, 0.3]$.

Let $\mathbf{x} := [\underline{x}, \overline{x}]$, where $\underline{x} \geq 0$, $\overline{x} \geq 0$ $(\underline{x}, \overline{x} \in \mathbb{R}^+)$, and $\mathbf{y} := [\underline{y}, \overline{y}]$, where $\underline{y} \geq 0$, $\overline{y} \geq 0$ $(\underline{y}, \overline{y} \in \mathbb{R}^+)$, be two non-negative interval variables. Let $\mathbf{f}(t) = [f(\underline{t}), f(\overline{t})]$ be a generalized

interval function, where $\mathbf{t} = [\underline{t}, \overline{t}]$ is an interval variable. The arithmetic operations of generalized intervals based on the Kaucher arithmetic are defined as follows:

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \overline{x} + \overline{y}], \tag{1}$$

$$x - dual\,\mathbf{y} = [\underline{x} - \underline{y}, \overline{x} - \overline{y}], \tag{2}$$

$$\mathbf{x} \times \mathbf{y} = [\underline{x} \times \underline{y}, \overline{x} \times \overline{y}], \tag{3}$$

$$\mathbf{x}/dual\,\mathbf{y} = [\underline{x}/\underline{y}, \overline{x}/\overline{y}], \underline{y} \neq 0, \overline{y} \neq 0 \tag{4}$$

$$log\,\mathbf{x} = [log\,\underline{x}, log\,\overline{x}], \underline{x} \neq 0, \overline{x} \neq 0 \tag{5}$$

Note that the boldface symbols represent generalized intervals in this paper. The greater than or equal to partial order relationship between two generalized intervals is defined as

$$[\underline{x}, \overline{x}] \geq [\underline{y}, \overline{y}] \Leftrightarrow \underline{x} \geq \underline{y} \wedge \overline{x} \geq \overline{y}. \tag{6}$$

2.2. Generalized interval probability

The generalized interval probability [5] is defined as follows. Given a sample space Ω and a σ-algebra A of random events over Ω, the generalized interval probability $\mathbf{p} \in \mathbb{KR}$ is defined as $\mathbf{p}: A \to [0,1] \times [0,1]$ which obeys the axioms of Kolmogorov: (1) $\mathbf{p}(\Omega) = [1, 1]$; (2) $[0,0] \leq \mathbf{p}(E) \leq [1, 1]$ $(\forall E \in A)$; and (3) for any countable mutually disjoint events $E_i \cap E_j = \Phi(i \neq j)$, $\mathbf{p}(\cup_{i=1}^n E_i) = \sum_{i=1}^n \mathbf{p}(E_i)$.

The most important property of the generalized interval probability is the *logic coherence constraint* (LCC): That is, for a mutually disjoint event partition $\cup_{i=1}^n E_i = \Omega$, $\sum_{i=1}^n \mathbf{p}(E_i) = 1$. The calculus structure of generalized interval probability is very similar to the one in the classical probability. The computation is greatly simplified compared to other interval probability representations such as the Dempster–Shafer evidence theory [12].

2.3. Generalized hidden Markov model

The GHMM is a generalization of HMM in the context of generalized interval probability theory. In GHMM, all probability values of HMM are replaced by generalized interval probabilities. A GHMM is defined as follows. The values of hidden states are in the form of $S = \{S_1, S_2, ..., S_N\}$, where N is the total number of possible hidden states. The hidden state variable at time t is $\mathbf{q}_t$, where $\mathbf{q}_t := [\underline{q}_t, \overline{q}_t]$. The M possible distinct observation symbols are $V = \{v_1, v_2, ..., v_M\}$. The generalized observation sequence is in the form of $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_T)$ where $\mathbf{o}_t$ is the observation value at time t. Note that the observations have the values of generalized intervals as random sets. Equivalently the lower and upper bounds can be viewed separately. $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ denotes the lower bound of the observation sequence, and $\overline{O} = (\overline{o}_1, \overline{o}_2, ..., \overline{o}_T)$ denotes the upper bound, where the value of $\underline{o}_t$ and $\overline{o}_t$ $(t = 1, ..., T)$ can be any of $\{v_1, v_2, ..., v_M\}$.

Let $q_t \in \text{pro}[\underline{q}_t, \overline{q}_t]$ and $o_t \in \text{pro}[\underline{o}_t, \overline{o}_t]$ be real-valued random variables that are included in the respective interval-valued random sets $[\underline{q}_t, \overline{q}_t]$ and $[\underline{o}_t, \overline{o}_t]$. $\mathbf{A} = (\mathbf{a}_{ij})_{N \times N}$ is the

state transition interval probability matrix, $\boldsymbol{a}_{ij} = \mathbf{p}(q_{t+1} = S_j | q_t = S_i)$, $(1 \leq i, j \leq N)$ is the interval probability of the transition from state S_i at time t to state S_j at time $t+1$. $\boldsymbol{B} = (\boldsymbol{b}_j(k))_{N \times M}$ is the observation interval probability matrix. $\boldsymbol{b}_j(k) = \mathbf{p}(o_t = v_k | q_t = S_j)$, $(1 \leq j \leq N, 1 \leq k \leq M)$ is the interval probability of observations in state S_j at time t. $\boldsymbol{\pi} = (\pi_i)_{1 \times N}$ is the initial state interval probability distribution, where $\pi_i = \mathbf{p}(q_1 = S_i)$, $(1 \leq i \leq N)$. The compact GHMM is denoted as $\lambda = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$.

Analogously to the classical HMM, the GHMM also have three basic problems to solve in real applications. The training problem of the GHMM is the crucial one. Its goal is to optimize the model parameters so that we can obtain the best model for the actual application scenario. A generalized Baum–Welch algorithm (GBWA) in the context of the generalized interval probability is proposed to provide an efficient approach to train GHMM. The GBWA is based on a generalized Jensen inequality, which is introduced in the following section.

3. Generalized Jensen inequality

The commonly used criteria for the optimization of HMM parameters include the maximum likelihood [13], the maximum mutual information [14], and the minimum discriminate information [15]. The maximum likelihood estimation by Baum–Welch algorithm [16] based on Jensen inequality is often adopted for optimizing the HMM parameters. Jensen inequality, named after Johan Jensen in 1906, relates the value of a convex function of an integral to the integral of the convex function [17]. As an important mathematical tool it has been widely used, such as for calculation of probability density function, statistical physics, information theory, and optimization. The use of Jensen inequality is based on the precise value. Here, interval values are used instead.

3.1. Generalized convex function

A generalized interval function $\mathbf{f}(\mathbf{t})$ is a generalized convex function if

$$\mathbf{f}(\mathbf{r}_1 \mathbf{t}_1 + \mathbf{r}_2 \mathbf{t}_2) \leq \mathbf{r}_1 \mathbf{f}(\mathbf{t}_1) + \mathbf{r}_2 \mathbf{f}(\mathbf{t}_2), \tag{7}$$

where $\mathbf{t} := [\underline{t}, \overline{t}]$, $\mathbf{r}_1 := [\underline{r}_1, \overline{r}_1]$, $\mathbf{r}_2 := [\underline{r}_2, \overline{r}_2]$, $\underline{t} \geq 0, \overline{t} \geq 0$, $\underline{r}_1 \geq 0, \overline{r}_1 \geq 0, \underline{r}_2 \geq 0, \overline{r}_2 \geq 0$, and $\mathbf{r}_1 + \mathbf{r}_2 = [1, 1]$. Under the same conditions, it is a generalized concave function if

$$\mathbf{f}(\mathbf{r}_1 \mathbf{t}_1 + \mathbf{r}_2 \mathbf{t}_2) \geq \mathbf{r}_1 \mathbf{f}(\mathbf{t}_1) + \mathbf{r}_2 \mathbf{f}(\mathbf{t}_2), \tag{8}$$

where $\mathbf{f}(t)$ is a generalized concave function. For instance, $log(\mathbf{t})$ defined in Eq. (5) is a generalized concave function.

3.2. Generalized Jensen inequality

For a generalized convex function $\mathbf{f}(t)$ and $\mathbf{r}_i := [\underline{r}_i, \overline{r}_i]$, $\underline{r}_i > 0, \overline{r}_i > 0 (i = 1, 2, ..., n), \sum_{i=1}^{n} \mathbf{r}_i = [1, 1]$, the generalized Jensen inequality can be stated as

$$\mathbf{f}\left(\sum_{i=1}^{n} \mathbf{r}_i \mathbf{t}_i\right) \leq \sum_{i=1}^{n} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) \tag{9}$$

The mathematical induction is adopted to prove Eq. (9). When $n=2$, Eq. (9) is defined in Eq. (7). Suppose $n=k$ and

$$\mathbf{f}\left(\sum_{i=1}^{k} \mathbf{r}_i \mathbf{t}_i\right) \leq \sum_{i=1}^{k} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i)$$

We need to prove that Eq. (9) is also correct when $n=k+1$.

$$\begin{aligned}
\mathbf{f}\left(\sum_{i=1}^{k+1} \mathbf{r}_i \mathbf{t}_i\right) &= \mathbf{f}\left(\sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{t}_i + (\mathbf{r}_k + \mathbf{r}_{k+1})\right. \\
&\quad \left. \times \left(\frac{\mathbf{r}_k}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_k + \frac{\mathbf{r}_{k+1}}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_{k+1}\right)\right) \\
&\leq \sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) + (\mathbf{r}_k + \mathbf{r}_{k+1}) \mathbf{f}\left(\frac{\mathbf{r}_k}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_k\right. \\
&\quad \left. + \frac{\mathbf{r}_{k+1}}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_{k+1}\right) \leq \sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) + \mathbf{r}_k \mathbf{f}(\mathbf{t}_k) \\
&\quad + \mathbf{r}_{k+1} \mathbf{f}(\mathbf{t}_{k+1}) \\
&= \sum_{i=1}^{k+1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i).
\end{aligned}$$

Thus, we can obtain Eq. (9) for all $i(i = 1, 2, ..., n)$.

It is straightforward that the opposite is true for a generalized concave function

$$\mathbf{f}\left(\sum_{i=1}^{n} \mathbf{r}_i \mathbf{t}_i\right) \geq \sum_{i=1}^{n} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) \tag{10}$$

The mathematical induction is also adopted to prove Eq. (10). When $n=2$, Eq. (10) is defined in Eq. (8). Suppose $n=k$ and $\mathbf{f}(\sum_{i=1}^{k} \mathbf{r}_i \mathbf{t}_i) \geq \sum_{i=1}^{k} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i)$. When $n=k+1$

$$\begin{aligned}
\mathbf{f}\left(\sum_{i=1}^{k+1} \mathbf{r}_i \mathbf{t}_i\right) &= \mathbf{f}\left(\sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{t}_i + (\mathbf{r}_k + \mathbf{r}_{k+1})\left(\frac{\mathbf{r}_k}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_k\right.\right. \\
&\quad \left.\left. + \frac{\mathbf{r}_{k+1}}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_{k+1}\right)\right) \\
&\geq \sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) + (\mathbf{r}_k + \mathbf{r}_{k+1}) \mathbf{f}\left(\frac{\mathbf{r}_k}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_k\right. \\
&\quad \left. + \frac{\mathbf{r}_{k+1}}{dual(\mathbf{r}_k + \mathbf{r}_{k+1})} \mathbf{t}_{k+1}\right) \geq \sum_{i=1}^{k-1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i) + \mathbf{r}_k \mathbf{f}(\mathbf{t}_k) \\
&\quad + \mathbf{r}_{k+1} \mathbf{f}(\mathbf{t}_{k+1}) = \sum_{i=1}^{k+1} \mathbf{r}_i \mathbf{f}(\mathbf{t}_i).
\end{aligned}$$

Thus, we can obtain that Eq. (10) is correct for all $i(i = 1, 2, ..., n)$.

4. Optimization methods in training process of the GHMM

During the training of GHMM, given the generalized observation sequence $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, ..., \boldsymbol{o}_T)$, the model parameters $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$ are adjusted to maximize the lower and the upper bounds of $\mathbf{p}(\boldsymbol{O}|\lambda)$. Different from the training of HMM, we can choose $\tilde{\lambda} = \{\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\pi}}\}$ so that the lower and the upper bounds of $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda})$ are both locally maximized by using an iterative procedure of the generalized Baum–Welch algorithm. The optimization method is based on the following two inferences that are similar to the ones in the HMM re-estimation [18].

Let $\boldsymbol{u}_i := [\underline{u}_i, \overline{u}_i]$, $\underline{u}_i > 0, \overline{u}_i > 0$, $i = 1, ..., S$ be positive interval numbers, and let $\boldsymbol{v}_i := [\underline{v}_i, \overline{v}_i]$, $\underline{v}_i \geq 0, \overline{v}_i \geq 0, i = 1, ..., S$ be nonnegative interval numbers such that

$\sum_i \boldsymbol{v}_i \geq [0,0]$. Then according to the generalized concavity of the *log* function and the generalized Jensen inequality,

$$
\begin{aligned}
\log \frac{\sum_i \boldsymbol{v}_i}{dual \sum_i \boldsymbol{u}_i} &= \log \left[\sum_i \left(\left(\frac{\boldsymbol{u}_i}{dual \sum_k \boldsymbol{u}_k} \right) \frac{\boldsymbol{v}_i}{dual \boldsymbol{u}_i} \right) \right] \\
&\geq \sum_i \left(\frac{\boldsymbol{u}_i}{dual \sum_k \boldsymbol{u}_k} \log \frac{\boldsymbol{v}_i}{dual \boldsymbol{u}_i} \right) \\
&= \frac{1}{dual \sum_k \boldsymbol{u}_k} \left[\sum_i (\boldsymbol{u}_i \log \boldsymbol{v}_i - dual \boldsymbol{u}_i \log \boldsymbol{u}_i) \right]
\end{aligned}
\tag{11}
$$

Here every summation is from 1 to S.

If $\mathbf{c}_i = [\underline{c}_i, \overline{c}_i] > 0$ $(i = 1, ..., N)$, the maximum value of the generalized interval function

$$
\mathbf{f}(\mathbf{t}_1, ..., \mathbf{t}_N) = \sum_{i=1}^{N} \mathbf{c}_i \log \mathbf{t}_i
\tag{12}
$$

subject to constraint $\sum_{i=1}^{N} \mathbf{t}_i = 1$ is reached when

$$
\mathbf{t}_i = \mathbf{c}_i / dual \sum_{i=1}^{N} \mathbf{c}_i \quad (i = 1, ..., N)
\tag{13}
$$

This is obtained by setting the first derivatives of the modified lower and upper objective functions as

$$
\frac{\partial}{\partial \underline{t}_i} f(\underline{t}_1, ..., \underline{t}_N) + \underline{\mu} (\sum_i \underline{t}_i - 1) = 0,
\tag{14}
$$

and

$$
\frac{\partial}{\partial \overline{t}_i} f(\overline{t}_1, ..., \overline{t}_N) + \overline{\mu} (\sum_i \overline{t}_i - 1) = 0,
\tag{15}
$$

where the Lagrange multipliers are $\underline{\mu} = \sum_i^N \underline{c}_i$ and $\overline{\mu} = \sum_i^N \overline{c}_i$.

4.1. Auxiliary interval function

Let S be the number of the state sequences with length T. The generalized observations $O = (o_1, o_2, ..., o_T)$ are divided into the lower bound observation sequence $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ and the upper bound observation sequence $\overline{O} = (\overline{o}_1, \overline{o}_2, ..., \overline{o}_T)$. In order to illustrate the training of the GHMM, the lower bound observation sequence $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ is used as an example. Let $\boldsymbol{u}_s^l$ be the joint interval probability $\boldsymbol{u}_s^l := \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \lambda)$ given model λ where $\boldsymbol{Q}_s = (\boldsymbol{q}_{s,1}, \boldsymbol{q}_{s,2}, ..., \boldsymbol{q}_{s,T})$, and let $\boldsymbol{v}_s^l$ be the joint interval probability $\boldsymbol{v}_s^l := \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \tilde{\lambda})$ given model $\tilde{\lambda}$. Then we have

$$
\sum_s \boldsymbol{u}_s^l = \mathbf{p}(\underline{O} | \lambda), \quad \sum_s \boldsymbol{v}_s^l = \mathbf{p}(\underline{O} | \tilde{\lambda})
\tag{16}
$$

Let the lower bound of auxiliary interval function $\boldsymbol{H}^l(\lambda, \tilde{\lambda})$ be

$$
\boldsymbol{H}^l(\lambda, \tilde{\lambda}) = \sum_s \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \lambda) \log \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \tilde{\lambda})
\tag{17}
$$

By substituting Eq. (16) into Eq. (11), we have

$$
\log \frac{\mathbf{p}(\underline{O} | \tilde{\lambda})}{dual \mathbf{p}(\underline{O} | \lambda)} \geq \frac{1}{dual \mathbf{p}(\underline{O} | \lambda)} \left[\boldsymbol{H}^l(\lambda, \tilde{\lambda}) - dual \boldsymbol{H}^l(\lambda, \lambda) \right]
\tag{18}
$$

In Eq. (18), we can obtain $\mathbf{p}(\underline{O} | \tilde{\lambda}) \geq \mathbf{p}(\underline{O} | \lambda)$ if $\boldsymbol{H}^l(\lambda, \tilde{\lambda}) \geq \boldsymbol{H}^l(\lambda, \lambda)$. That is, if we can find a model $\tilde{\lambda}$ that makes the right-hand side of Eq. (18) positive, the model λ can be improved. Clearly, the largest guaranteed improvement by this method is to maximize $\boldsymbol{H}^l(\lambda, \tilde{\lambda})$, hence the maximum lower and upper bounds of $\mathbf{p}(\underline{O} | \tilde{\lambda})$ are obtained. The

maximum value of $\mathbf{p}(\overline{O} | \tilde{\lambda})$ for the upper bound observation sequence can be obtained similarly.

4.2. Training of the lower and the upper bounds

Different from Baum–Welch algorithm in HMM [18,19], the training procedure uses

$$
\begin{aligned}
\log \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \tilde{\lambda}) &= \log(\mathbf{p}(\boldsymbol{Q}_s | \tilde{\lambda}) \cdot \mathbf{p}(\underline{O} | \boldsymbol{Q}_s, \tilde{\lambda})) = \log \tilde{\boldsymbol{\pi}}_{\boldsymbol{q}_1}^l \\
&\quad + \sum_{t=1}^{T-1} \log \tilde{\boldsymbol{a}}_{\boldsymbol{q}_t \boldsymbol{q}_{t+1}}^l + \sum_{t=1}^{T} \log \tilde{\boldsymbol{b}}_{\boldsymbol{q}_t}^l(\underline{o}_t)
\end{aligned}
\tag{19}
$$

where $\tilde{\boldsymbol{a}}_{ij}^l, \tilde{\boldsymbol{b}}_j^l(k)$ and $\tilde{\boldsymbol{\pi}}_i^l$ are respectively the state transition probability, the observation probability, and the initial state probability corresponding to the lower bound observation sequence in the form of the generalized interval. Substituting Eq. (19) into Eq. (17), and re-grouping terms in the summation according to state transitions and observations, we have

$$
\begin{aligned}
\boldsymbol{H}^l(\lambda, \tilde{\lambda}) &= \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{c}_{ij}^l \log \tilde{\boldsymbol{a}}_{ij}^l + \sum_{j=1}^{N} \sum_{k=1}^{M} \boldsymbol{d}_{jk}^l \log \tilde{\boldsymbol{b}}_j^l(k) \\
&\quad + \sum_{i=1}^{N} \boldsymbol{e}_1^l \log \tilde{\boldsymbol{\pi}}_i^l
\end{aligned}
\tag{20}
$$

where

$$
\boldsymbol{c}_{ij}^l = \sum_s \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \lambda) \sum_{t=1}^{T-1} \xi_t^l(i,j) = \mathbf{p}(\underline{O} | \lambda) \sum_{t=1}^{T-1} \xi_t^l(i,j)
\tag{21}
$$

$$
\boldsymbol{d}_{jk}^l = \sum_s \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \lambda) \sum_{t=1, \underline{o}_t = v_k}^{T} \gamma_t^l(j) = \mathbf{p}(\underline{O} | \lambda) \sum_{t=1, \underline{o}_t = v_k}^{T} \gamma_t^l(j)
\tag{22}
$$

$$
\boldsymbol{e}_1^l = \sum_s \mathbf{p}(\underline{O}, \boldsymbol{Q}_s | \lambda) \gamma_1^l(i) = \mathbf{p}(\underline{O} | \lambda) \gamma_1^l(i)
\tag{23}
$$

where $\xi_t^l(i,j) = \mathbf{p}(q_t = S_i, q_{t+1} = S_j | \underline{O}, \lambda)$ is the lower interval probability of being in state S_i at time t and in state S_j at time $t+1$, and $\gamma_t^l(i) = \mathbf{p}(q_t = S_i | \underline{O}, \lambda)$ is the lower interval probability of being in state S_i at time t when the observation sequence $\underline{O}$ and the model λ are given. Thus, $\boldsymbol{c}_{ij}^l, \boldsymbol{d}_{jk}^l$ and $\boldsymbol{e}_1^l$ are the expected values of $\sum_{t=1}^{T-1} \xi_t^l(i,j)$, $\sum_{t=1, \underline{o}_t = v_k}^{T} \gamma_t^l(j)$, and $\gamma_1^l(i)$, respectively, based on model λ.

According to Eq. (13), $\boldsymbol{H}^l(\lambda, \tilde{\lambda})$ is maximized if and only if

$$
\tilde{\boldsymbol{a}}_{ij}^l = \frac{\boldsymbol{c}_{ij}^l}{dual \sum_j \boldsymbol{c}_{ij}^l} = \sum_{t=1}^{T-1} \xi_t^l(i,j) \bigg/ dual \sum_{t=1}^{T-1} \gamma_t^l(i)
\tag{24}
$$

$$
\tilde{\boldsymbol{b}}_j^l(k) = \boldsymbol{d}_{jk}^l / dual \sum_k \boldsymbol{d}_{jk}^l = \sum_{t=1, \underline{o}_t = v_k}^{T} \gamma_t^l(j) \bigg/ dual \sum_{t=1}^{T} \gamma_t^l(j)
\tag{25}
$$

$$
\tilde{\boldsymbol{\pi}}_i^l = \boldsymbol{e}_1^l / dual \sum_i \boldsymbol{e}_1^l = \gamma_1^l(i)
\tag{26}
$$

Eqs. (24)–(26) are regarded as the lower bound re-estimation formulas. The maximum value of $\boldsymbol{H}^l(\lambda, \tilde{\lambda})$ can be reached by the lower bound re-estimation formulas. Hence the maximum value of $\mathbf{p}(\underline{O} | \tilde{\lambda})$ is also obtained.

In a similar vein, we can obtain the upper bound re-estimation formulas as

$$
\tilde{\boldsymbol{a}}_{ij}^u = \boldsymbol{c}_{ij}^u / dual \sum_j \boldsymbol{c}_{ij}^u = \sum_{t=1}^{T-1} \xi_t^u(i,j) \bigg/ dual \sum_{t=1}^{T-1} \gamma_t^u(i)
\tag{27}
$$

$$\tilde{\boldsymbol{b}}_j^u(k) = \boldsymbol{d}_{jk}^u / dual\sum_k \boldsymbol{d}_{jk}^u = \sum_{t=1, \underline{o}_t=\nu_k}^T \gamma_t^u(j) \bigg/ dual\sum_{t=1}^T \gamma_t^u(j) \quad (28)$$

$$\tilde{\boldsymbol{\pi}}_i^u = \boldsymbol{e}_1^u / dual\sum_i \boldsymbol{e}_1^u = \gamma_1^u(i) \quad (29)$$

where $\tilde{\boldsymbol{a}}_{ij}^u$, $\tilde{\boldsymbol{b}}_j^u(k)$ and $\tilde{\boldsymbol{\pi}}_i^u$ are respectively the upper state transition probability, the upper observation probability, and the upper initial state probability used in the form of the generalized interval. $\xi_t^u(i,j)$ is the upper interval probability of being in state S_i at time t and in state S_j at time $t+1$, and $\gamma_t^u(i)$ is the upper interval probability of being in state S_i at time t when the observation sequence $\overline{O}$ and the model λ are given. The maximum values of $\boldsymbol{H}^u(\lambda, \tilde{\lambda})$ and $\mathbf{p}(\overline{O}|\tilde{\lambda})$ then can be obtained.

4.3. Training of the GHMM

According to the concept of multiple observation sequences [9], $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ and $\overline{O} = (\overline{o}_1, \overline{o}_2, ..., \overline{o}_T)$ are regarded as two independent observation sequences. The GHMM re-estimation formulas then are defined according to the EM algorithm of statistics as

$$\tilde{\boldsymbol{a}}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t^l(i,j) + \sum_{t=1}^{T-1} \xi_t^u(i,j)}{dual\left(\sum_{t=1}^{T-1} \gamma_t^l(i) + \sum_{t=1}^{T-1} \gamma_t^u(i)\right)} \quad (30)$$

$$\tilde{\boldsymbol{b}}_j(k) = \frac{\sum_{t=1, \underline{o}_t=\nu_k}^T \gamma_t^l(j) + \sum_{t=1, \underline{o}_t=\nu_k}^T \gamma_t^u(j)}{dual\left(\sum_{t=1}^T \gamma_t^l(j) + \sum_{t=1}^T \gamma_t^u(j)\right)} \quad (31)$$

$$\tilde{\boldsymbol{\pi}}_i = \tfrac{1}{2}(\gamma_1^l(i) + \gamma_1^u(i)) \quad (32)$$

where $\tilde{\boldsymbol{a}}_{ij}$, $\tilde{\boldsymbol{b}}_j(k)$ and $\tilde{\boldsymbol{\pi}}_i$ are the state transition interval probability, the observation interval probability, and the initial state interval probability, respectively [1,8]. Different from HMM [1], the re-estimation is based on

backward interval variable, and $\beta_t^u(i)$ are the upper backward interval variable, respectively defined as

$$\alpha_t^l(i) := \mathbf{p}(\underline{o}_1, \underline{o}_2, ..., \underline{o}_t, q_t=S_i|\lambda) \quad (36)$$

$$\alpha_t^u(i) := \mathbf{p}(\overline{o}_1, \overline{o}_2, ..., \overline{o}_t, q_t=S_i|\lambda) \quad (37)$$

$$\beta_t^l(i) := \mathbf{p}(\underline{o}_{t+1}, \underline{o}_{t+2}, ..., \underline{o}_T | q_t=S_i, \lambda) \quad (38)$$

$$\beta_t^u(i) := \mathbf{p}(\overline{o}_{t+1}, \overline{o}_{t+2}, ..., \overline{o}_T | q_t=S_i, \lambda) \quad (39)$$

The forward and the backward interval variables can be derived similar to the ones in HMM.

The trained model parameters $\tilde{\lambda} = \{\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\pi}}\}$ can be obtained by applying the re-estimation in Eqs. (33)–(35). With $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ and $\overline{O} = (\overline{o}_1, \overline{o}_2, ..., \overline{o}_T)$ regarded as two independent observation sequences, we can define

$$\mathbf{p}(\boldsymbol{O}|\lambda) := \mathbf{p}(\underline{O}|\lambda)\mathbf{p}(\overline{O}|\lambda) \quad (40)$$

$$\mathbf{p}(\boldsymbol{O}|\tilde{\lambda}) := \mathbf{p}(\underline{O}|\tilde{\lambda})\mathbf{p}(\overline{O}|\tilde{\lambda}) \quad (41)$$

where $\mathbf{p}(\boldsymbol{O}|\lambda)$ is the interval probability of generalized observation under the condition of initial model λ; $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda})$ is the interval probability of generalized observation under the condition of initial model $\tilde{\lambda}$. According to the lower and the upper bound re-estimation formulas, $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda}) \geq \mathbf{p}(\boldsymbol{O}|\lambda)$ can also be obtained, since the values of interval probabilities are between 0 and 1. In an iterative procedure, the value of $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda})$ is gradually increased. The final result $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda})$ is a so-called maximum likelihood estimation of GHMM.

The local maxima of GHMM parameters can be obtained by the iterative training algorithm as follows:

(1) Choose an initial model $\lambda = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$.
(2) Choose the generalized observation sequence $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, ..., \boldsymbol{o}_T)$, i.e. the lower bound $\underline{O} = (\underline{o}_1, \underline{o}_2, ..., \underline{o}_T)$ and the upper bound $\overline{O} = (\overline{o}_1, \overline{o}_2, ..., \overline{o}_T)$.
(3) Obtain the trained model $\tilde{\lambda} = \{\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\pi}}\}$ based on Eqs. (33)–(35).

$$\tilde{\boldsymbol{a}}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{dual\sum_{t=1}^{T-1} \gamma_t} = \frac{\sum_{t=1}^{T-1} \alpha_t^l(i)\boldsymbol{a}_{ij}\boldsymbol{b}_j(\underline{o}_{t+1})\beta_{t+1}^l(j)/dual\mathbf{p}(\underline{O}|\lambda) + \sum_{t=1}^{T-1} \alpha_t^l(i)\boldsymbol{a}_{ij}\boldsymbol{b}_j(\overline{o}_{t+1})\beta_{t+1}^l(j)/dual\mathbf{p}(\overline{O}|\lambda)}{dual\left(\sum_{t=1}^{T-1} \alpha_t^l(i)\beta_t^l(i)/dual\mathbf{p}(\underline{O}|\lambda) + \sum_{t=1}^{T-1} \alpha_t^l(i)\beta_t^l(i)/dual\mathbf{p}(\overline{O}|\lambda)\right)} \quad (1 \leq i, j \leq N), \quad (33)$$

$$\tilde{\boldsymbol{b}}_j(o_t=\nu_k) = \frac{\sum_{t=1, \boldsymbol{o}_t=\nu_k}^T \gamma_t(j)}{dual\sum_{t=1}^T \gamma_t(j)}$$

$$= \frac{\sum_{t=1, o_t=\nu_k}^T \left(\alpha_t^l(j)\beta_t^l(j)/dual\mathbf{p}(\underline{O}|\lambda) + \alpha_t^u(j)\beta_t^u(j)/dual\mathbf{p}(\overline{O}|\lambda)\right)}{dual\sum_{t=1}^T \left(\alpha_t^l(j)\beta_t^l(j)/dual\mathbf{p}(\underline{O}|\lambda) + \alpha_t^u(j)\beta_t^u(j)/dual\mathbf{p}(\overline{O}|\lambda)\right)}$$

$$(1 \leq j \leq N, 1 \leq k \leq M), \quad (34)$$

$$\tilde{\boldsymbol{\pi}}_i = \frac{1}{2}\left(\frac{\alpha_1^l(i)\beta_1^l(i)}{dual\mathbf{p}(\underline{O}|\lambda)} + \frac{\alpha_1^u(i)\beta_1^u(i)}{dual\mathbf{p}(\overline{O}|\lambda)}\right) \quad (1 \leq i \leq N), \quad (35)$$

where $\alpha_t^l(i)$ is the lower forward interval variable, $\alpha_t^u(i)$ is the upper forward interval variable, $\beta_t^l(i)$ is the lower

(4) If the local or global optimum of $\mathbf{p}(\boldsymbol{O}|\tilde{\lambda})$ is reached, then stop; otherwise, go back to the training model and use $\tilde{\lambda}$ to replace λ.

5. Case studies

5.1. Recognition of tool wear

5.1.1. Experimental setup and data processing

In order to demonstrate the performance of the proposed optimization method for training GHMM, a training model of tool wear is developed. The experimental bench is illustrated in Fig. 1. The cutting tests are conducted on Mikron UCP800 Duro, which is a five-axis machining center. The thrust force is measured by a Kistler 9253823

dynamometer. The resulted signals are converted into output voltages. Then these voltage signals are amplified by Kistler multichannel charge amplifier 5070. Force signals are simultaneously recorded by NI PXIe-1802 data recorder with 5 kHz sampling frequency. The computer screen is connected by a junction box NI SCB-68. The collected signals are displayed by Cathode ray tube CRT. The workpiece is 300M steel material by heat treatment. A carbide cutter SANDVIK R216.34-20050, IAK38H 1620, ⌀20.0 mm is selected as a cutting tool. The cutting tool is zoomed in by Olympus STM6 measure microscope and recorded by a computer. The workpiece is continuously processed under different processing conditions until the obvious cutting tool wear is observed.

The tool states are classified into three categories: the initial processing status of the tool is named *sharp* state (pattern 1), the wear processing status of the tool is named *wear* state (pattern 3), and the status between sharp state and wear state is named *slight wear* state (pattern 2) [20]. The pictures of sharp cutting tool I and wear cutting tool II, with a 570× magnification under the same calibration condition, are shown in Fig. 2.

The real-time cutting processing signals under a sharp cutting tool condition and a wear cutting tool condition are shown as Fig. 3. Signal I represents the sharp cutting tool condition. Its processing condition is defined by a 1000 rpm spindle speed, a 400 mm/min feed rate, a 2 mm cutting depth, and a 2 mm cutting width. Signal II represents the wear cutting tool condition. Its processing con-

dition is defined by a 2000 rpm spindle speed, a 70 mm/min feed rate, a 5 mm cutting depth, and a 1 mm cutting width. We can see that the amplitude increases significantly under the sharp cutting condition in Fig. 3.

The fast Fourier transform (FFT) processing results of the time domain signals in the sharp cutting tool and wear cutting tool conditions are shown in Fig. 4. It shows that the frequency spectrum is significantly different under different cutting tool conditions.

Wavelet analysis, which can record the detailed information of the different frequency band, is a superior time–frequency analysis tool [21]. In this work, the four-level wavelet packet decomposition is used. The root mean square (RMS) of the wavelet coefficients at different scales is shown in Fig. 5. It can be found that RMS results are significantly different for the three states.

To quantify measurement errors as the source of epistemic uncertainty, each value of experimental data is converted into a form of generalized interval by considering an error of ±5%. Then the four-level wavelet packet decomposition is applied. The RMS of the wavelet coefficients at different scales are taken as the feature observations vector. The training procedure for finding the optimal model is carried out. The GHMM convergence curve of log-likelihood is shown in Fig. 6. It can be seen that the GHMM training process has a good convergence property. The optimized GHMM can be obtained.

5.1.2. Recognition of tool wear

The state recognition of tool wear plays a critical role in the automation processing of machine tools. However, the
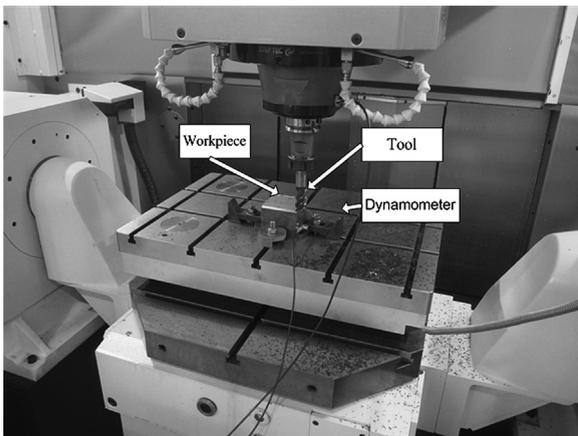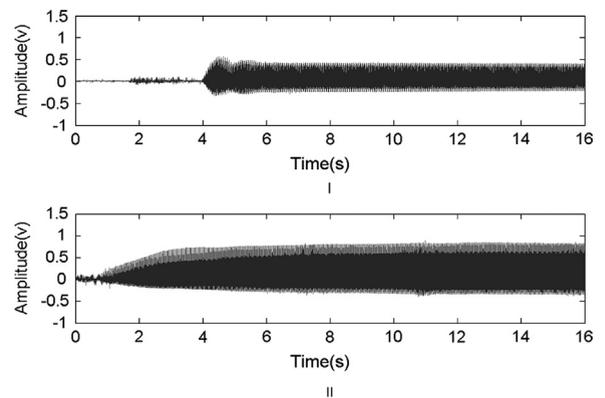

Fig. 1. Experimental bench for cutting processing.
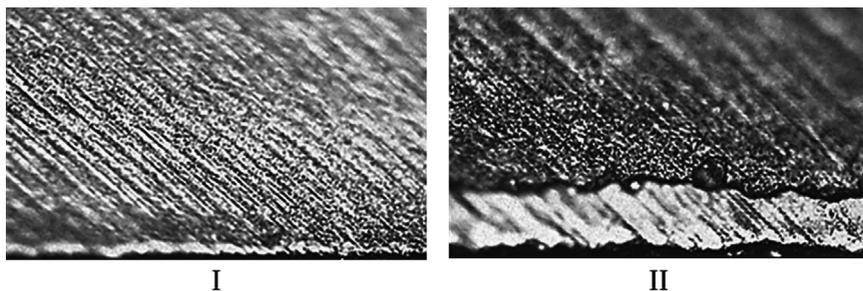

Fig. 3. Dynamometer signals.

Fig. 2. The pictures of cutting edge: sharp cutting edge (I), wear cutting edge (II).
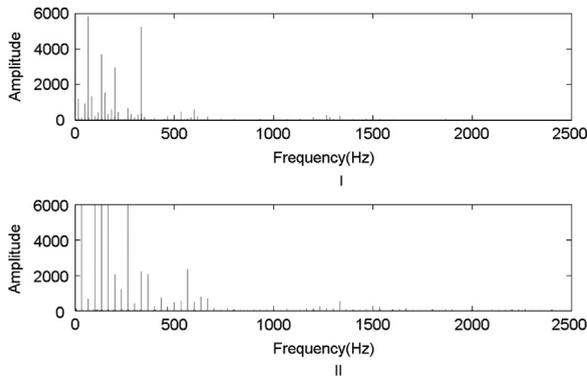
Fig. 4. The frequency spectrum: sharp cutting state (I) and wear cutting state (II).
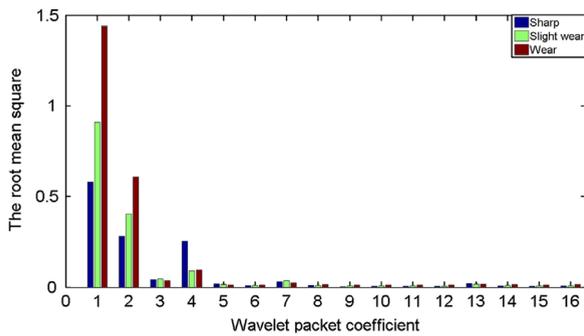

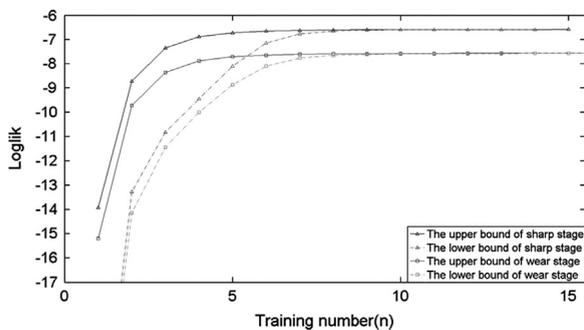**Fig. 5.** The RMS of coefficient in three states.

Fig. 6. The training convergence curve.

state recognition is not an easy task for some reasons. For instance, the machining processes are non-linear and time varying, which makes it difficult to model. The acquired sensor signals are affected by some uncertainties such as geometry variances, workpiece material properties, digitizer's noise, and sensor nonlinearity. Recently some classification methods are used for cutting state recognition. For example, the feature extraction of wavelet transform is used to monitor tool wear [22]; a HMM classifier is used as monitoring the cutting tool condition [23]; an artificial neural network is applied as an on-line and indirect tool for wear monitoring [24]; and a support vector machine is used for state recognition of tool wear [20]. Among these methods, epistemic uncertainty is often ignored.
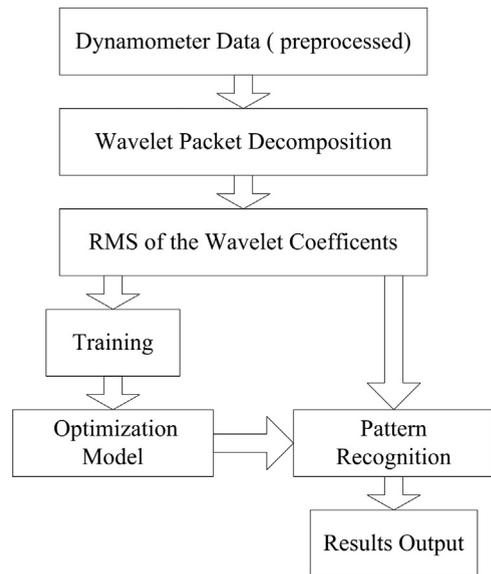

Fig. 7. Flow chart of the tool states recognition.

In this paper, the GHMM enhances the reasoning process. As shown in Fig. 7, a system for tool state recognition based on GHMM is designed. It is composed of the wavelet-based feature extraction and the RMS of the wavelet coefficients for GHMM input. Each GHMM pattern is trained by the RMS of lower and upper bounds from post treatment, and test sample is recognized by the GHMM based classification method.

In this system, 10 signals of each pattern are taken out from 60 signals for training. Each pattern of the optimized GHMM with respect to the corresponding state is established, and the remaining signals of each pattern are used to test the validity of the models. For each optimized GHMM, the current observation sequence of the test sample is substituted. The log-likelihood values related to three optimized GHMMs are calculated. The log-likelihoods of test samples with respect to the optimized models are compared. The output states with the maximum log-likelihoods are selected. Here, the maxi–min criterion (pessimistic criterion) [25] of interval comparison is adopted to improve the reliability of estimation. In this criterion, the minimal result for each interval will be chosen first. Then the maximal value of these minimal results is selected. The GHMM recognition results of tool wear are shown in Table 1. The results are compared with the ones by the same recognition procedure based on the traditional HMM [26].

5.1.3. Result analysis

As shown in Table 1, most samples have been recognized correctly, where accuracy rate of GHMM (95%) is higher than accuracy rate of HMM (91.7%). In the GHMM-based classification method, the three incorrect recognition results are consistent with the tendency of tool wear, and it will not damage the cutting tools. However, in the HMM-based classification method, there are two incorrect recognition results (nos.43 and 53) where a wear state is mistakenly recognized as a sharp state. As a result, replacement of a worn tool will be postponed and cause

Table 1
Pattern classification results of the tool wear recognition.

No.	Cutting depth (mm)	Cutting width (mm)	Spindle speed (rpm)	Feed rate (mm/min)	Test pattern	Output of GHMM	Output of HMM	Recognition result of GHMM	Recognition result of HMM
1	0.5	2	1000	240	1	1	1	Correct	Correct
2	0.5	2	1000	280	1	1	1	Correct	Correct
3	0.5	2	1000	320	1	1	1	Correct	Correct
4	0.5	2	1000	360	1	1	1	Correct	Correct
5	0.5	2	1000	400	1	1	1	Correct	Correct
6	0.5	3	1000	200	1	1	1	Correct	Correct
7	0.5	3	1000	240	1	1	1	Correct	Correct
8	0.5	3	1000	280	1	1	1	Correct	Correct
9	0.5	3	1000	320	1	1	1	Correct	Correct
10	0.5	3	1000	360	1	1	1	Correct	Correct
11	0.5	3	1000	400	1	1	1	Correct	Correct
12	0.5	4	1000	200	1	1	1	Correct	Correct
13	0.5	4	1000	240	1	1	1	Correct	Correct
14	0.5	4	1000	280	1	1	2	Correct	Incorrect
15	0.5	4	1000	320	1	1	1	Correct	Correct
16	0.5	4	1000	360	1	1	1	Correct	Correct
17	0.5	4	1000	400	1	2	1	Incorrect	Correct
18	0.5	5	1000	200	1	1	1	Correct	Correct
19	0.5	5	1000	240	1	1	1	Correct	Correct
20	0.5	5	1000	280	1	1	1	Correct	Correct
21	0.5	5	1000	320	1	1	1	Correct	Correct
22	1	2	1000	240	1	1	1	Correct	Correct
23	1	2	1000	320	1	2	2	Incorrect	Incorrect
24	1.5	2	1000	200	1	1	1	Correct	Correct
25	3	0.6	1000	320	2	2	2	Correct	Correct
26	3	0.6	1000	360	2	2	2	Correct	Correct
27	3	0.6	1000	400	2	2	2	Correct	Correct
28	3	0.8	1000	200	2	2	2	Correct	Correct
29	3	0.8	1000	240	2	2	2	Correct	Correct
30	3	0.8	1000	280	2	2	2	Correct	Correct
31	3	0.8	1000	320	2	2	2	Correct	Correct
32	6	0.4	1000	320	2	2	2	Correct	Correct
33	6	0.4	1000	360	2	2	2	Correct	Correct
34	6	0.4	1000	400	2	2	2	Correct	Correct
35	6	0.6	1000	200	2	2	2	Correct	Correct
36	6	0.6	1000	240	2	2	2	Correct	Correct
37	6	0.6	1000	280	2	2	2	Correct	Correct
38	9	0.4	1000	280	2	2	2	Correct	Correct
39	9	0.4	1000	320	2	2	2	Correct	Correct
40	9	0.4	1000	360	2	2	2	Correct	Correct
41	9	0.6	1000	200	2	2	2	Correct	Correct
42	12	0.4	1000	320	2	3	3	Incorrect	Incorrect
43	5	1	2000	70	3	3	1	Correct	Incorrect
44	5	1	2000	70	3	3	3	Correct	Correct
45	5	1	2000	70	3	3	3	Correct	Correct
46	5	1	2000	70	3	3	3	Correct	Correct
47	5	1	2000	70	3	3	3	Correct	Correct
48	5	1	2000	70	3	3	3	Correct	Correct
49	5	1	2000	70	3	3	3	Correct	Correct
50	5	1	2000	70	3	3	3	Correct	Correct
51	5	1	2000	70	3	3	3	Correct	Correct
52	5	1	2000	70	3	3	3	Correct	Correct
53	5	1	2000	70	3	3	1	Correct	incorrect
54	5	1	2000	70	3	3	3	Correct	Correct
55	5	1	2000	70	3	3	3	Correct	Correct
56	5	1	2000	70	3	3	3	Correct	Correct
57	5	1	2000	70	3	3	3	Correct	Correct
58	5	1	2000	70	3	3	3	Correct	Correct
59	5	1	2000	70	3	3	3	Correct	Correct
60	5	1	2000	70	3	3	3	Correct	Correct

tool damage. The comparison shows that the GHMM-based classification method is superior to the HMM-based classification method.

Furthermore, in the proposed GHMM-based classification method, the log-likelihood values are in the form of generalized interval probability. The width of an interval

probability quantifies the extent of epistemic uncertainty component. The results can improve the reliability of recognition by using the extra information provided by the interval values. For example, three log-likelihood values of no. 17 in Table 1, with respect to the three optimized GHMMs, are shown in Table 2. We can find that interval $[-60.0859, -28.9406]$ contains interval $[-41.3431, -35.6949]$. The recognition result is sharp state, which is correct by the maxi-max criterion (optimistic criterion) of interval comparison [25]. This provides the extra information that the state could be possibly misinterpreted, and the fact may be also misinterpreted. Therefore, we are informed that it is better to conduct more experiments and collect more feature data so that more accurate decision can be made. In contrast, the HMM-based classification method by which a precise probability value does not provide such information. For example, three log-likelihood values of no. 14 in Table 1, with respect to the three optimized HMMs, are shown in Table 3. The recognition result is a slight wear state, which is incorrect by a simple comparison of the log-likelihood values.

5.2. Recognition of cutting states

A second example of GHMM application to the recognition of cutting states is described here. The cutting experimental setup is shown as Fig. 8. The experiments are conducted in a numerical control milling machine tool DM4600. The tool is a high-speed steel SNMG120412-MA UE6020 with 20 mm diameter, tool bar 80 mm length, and three teeth. The material of workpiece is aluminum alloy 6160, and the shape of workpiece is rectangular with a cut top right corner. The spindle speed is kept constant at 3400 rpm, and the feed rate is 1020 mm/min. 0.2 mm radial cutting depth and 12–25 mm axial cutting depth are adopted. The vibration signals of spindle are obtained by gradually increasing axial depth of cutting. Two accelerometers PCB-352C33 are settled on the spindle housing along the X and Y directions of the machine tool, respectively. A data recorder unit NI PXI-1042 with a 5120 Hz sampling rate is used to acquire acceleration signals.

Standard deviation (STD) [27], power spectrum density (PSD) [28], and mean square frequency (MSF) [29], as the features of pattern recognition, are used to detect cutting states. To quantify measurement errors, these feature values are considered to have a general error $\pm 5\%$. The cutting states are divided into three stages: stable stage I, transitional stage II, and chatter stage III. The interval forms of STD, PSD and MSF are considered as multiple observation sequences in the proposed GHMM. Four signal groups of each pattern are taken out of the 33 signal groups to train GHMM. All signals

of each pattern are used for test. The test results are shown in Table 4. It is shown that the success rate of the cutting state recognition is 100%. With the number of feature numbers increased, the success rate may be increased by the analysis of experimental results.

6. Conclusion

The need to consider aleatory and epistemic uncertainties has been widely recognized. In this paper an optimization method GBWA based on a generalized interval probability theory to distinguish the two uncertainty components is proposed. The observation sequence is viewed separately as the lower and the upper bound observation sequences. A generalized Baum–Welch′s auxiliary function and a generalized Jensen inequality are used. Similar to HMM training, a set of training equations are derived. The lower and upper bound re-estimation formulas have been developed based on a multiple observation concept. According to the multiple observations EM algorithm, this method guarantees the local maxima for the lower and upper bound observation sequences. Two cases of tool state

Table 3
The recognition result of no. 14 based on the HMM.

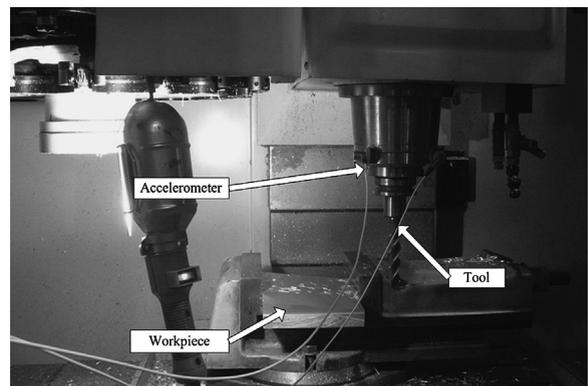| Test sample	Log-likelihood values Optimized models		
	Sharp optimized model	Slight wear optimized model	Wear optimized model
No. 17 (Sharp)	-12.6576	-11.6025	-60.0658

Fig. 8. Experimental setup for milling processing.

Table 2
The recognition result of no. 17 based on the GHMM.

Test sample	Log-likelihood values Optimized models		
	Sharp optimized model	Slight wear optimized model	Wear optimized model
No. 17 (Sharp)	$[-60.0859, -28.9406]$	$[-41.3431, -35.6949]$	$[-73.9063, -171.8820]$

Table 4
Pattern classification results of cutting states.

No.	Radial depth of cutting (mm)	Axial depth of cutting(mm)	Spindle speed (rpm)	Feed rate (mm/min)	Test pattern	Output of GHMM	Recognition result of GHMM
1	0.2	12	3400	1020	I	I	Correct
2	0.2	12	3400	1020	II	II	Correct
3	0.2	12	3400	1020	III	III	Correct
4	0.2	13	3400	1020	I	I	Correct
5	0.2	13	3400	1020	II	II	Correct
6	0.2	13	3400	1020	III	III	Correct
7	0.2	14	3400	1020	I	I	Correct
8	0.2	14	3400	1020	II	II	Correct
9	0.2	14	3400	1020	III	III	Correct
10	0.2	16	3400	1020	I	I	Correct
11	0.2	16	3400	1020	II	II	Correct
12	0.2	16	3400	1020	III	III	Correct
13	0.2	15	3400	1020	I	I	Correct
14	0.2	15	3400	1020	II	II	Correct
15	0.2	15	3400	1020	III	III	Correct
16	0.2	17	3400	1020	I	I	Correct
17	0.2	17	3400	1020	II	II	Correct
18	0.2	17	3400	1020	III	III	Correct
19	0.2	18	3400	1020	I	I	Correct
20	0.2	18	3400	1020	II	II	Correct
21	0.2	18	3400	1020	III	III	Correct
22	0.2	21	3400	1020	I	I	Correct
23	0.2	21	3400	1020	II	II	Correct
24	0.2	21	3400	1020	III	III	Correct
25	0.2	22	3400	1020	I	I	Correct
26	0.2	22	3400	1020	II	II	Correct
27	0.2	22	3400	1020	III	III	Correct
28	0.2	24	3400	1020	I	I	Correct
29	0.2	24	3400	1020	II	II	Correct
30	0.2	24	3400	1020	III	III	Correct
31	0.2	25	3400	1020	I	I	Correct
32	0.2	25	3400	1020	II	II	Correct
33	0.2	25	3400	1020	III	III	Correct

and cutting state recognition in manufacturing processes are used to demonstrate the new method. The results show that the GHMM based classification has a good recognition performance. With the two kinds of uncertainty quantified by the generalized interval probability, the GHMM-based recognition results provide more information, thus more robust decisions can be made. Yet there are some limitations for the proposed approach. The computational cost of the GHMM is higher than that of the classical HMM, since the data of GHMM observation sequences are twice as much as that of HMM. How to reduce the computational cost of GHMM will be included in the future study.

Acknowledgments

References

[1] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286.

[2] N.U. Nair, T.V. Sreenivas, Multi-pattern Viterbi algorithm for joint decoding of multiple speech patterns, Signal Processing 90 (12) (2010) 3278–3283.

[3] G.B. Song, P. Martynovich, A study of hmm-based bandwidth extension of speech signals, Signal Processing 89 (10) (2009) 2036–2044.

[4] M. Dong, D. He, P. Banerjee, J. Keller, Equipment health diagnosis and prognosis using hidden semi-Markov models, International Journal of Advanced Manufacturing Technology 30 (7-8) (2006) 738–749.

[5] Y. Wang, Imprecise probabilities based on generalized intervals for system reliability assessment, International Journal of Reliability & Safety 4 (2010) 319–342.

[6] E. Kaucher, Interval analysis in the extended interval space IR, Computing Supplement 2 (1980) 33–49.

[7] Y. Wang, Multiscale uncertainty quantification based on a generalized hidden Markov model, ASME Journal of Mechanical Design 3 (2011) 1–10.

[8] Y.M. Hu, F.Y. Xie, B. Wu, Y. Cheng, G.F. Jia, Y. Wang, M.Y. Li, An optimization method for training generalized hidden Markov model based on generalized Jensen inequality, in: Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics, 2012, pp. 268–274, doi: 10.5220/0004118202680274.

[9] X.L. Li, M. Parizeau, R. Plamondon, Training hidden Markov models with multiple observations—a combinatorial method, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (4) (2000) 371–377.

[10] E.D. Popova, All about generalized interval distributive relations. I. Complete Proof of the Relations Sofia, 2000.

[11] E. Gardenes, M.A. Sainz, L. Jorba, R. Calm, R. Estela, H. Mielgo, A. Trepat, Modal intervals, Reliable Computing 2 (2001) 77–111.

[12] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, NJ, USA, 1976.

[13] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royial Statistical Society 39 (1977) 1–38.

[14] L. Bahl, P. Brown, P. Souza, R. Mercer, Maximum mutual information estimation of hidden Markov model parameters for speech recognition, in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP, 1986, pp. 49–52.

[15] S. Kullback, M. Khairat, A note on minimum discrimination information, The Annals of Mathematical Statistics 37 (1965) 279–280.

[16] L.E. Baum, T. Petrie, G. Souls, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Annals of Mathematical Statististics 41 (1) (1970) 164–171.

[17] J. Jensen, Sur les functions convexes et les inégalités entre les valeurs moyennes, Acta Mathematica 30 (1) (1906) 175–193.

[18] S. Levinson, R. Rabiner, M. Sondhi, An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, Bell Systems Technical Journal 62 (1983) 1035–1074.

[19] B.H. Juang, L.R. Rabiner, Hidden Markov models for speech recognition, Technometrics 33 (3) (1991) 251–272.

[20] S. Guan, L.S. Wang, Study on identification method of tool wear based on singular value decomposition and least squares support vector machine, Advances in Intelligent and Soft Computing 112 (2012) 835–843.

[21] N. Fang, S.P. Pai, S Mosquea, Effect of tool edge wear on the cutting forces and vibrations in high-speed finish machining of Inconel 718: an experimental study and wavelet transform analysis, International Journal of Advanced Manufacturing Technology 52 (2011) 65–77.

[22] Y. Choi, R. Narayanaswami, A. Chandra, Tool wear monitoring in ramp cuts in end milling using the wavelet transform, International Journal of Advanced Manufacturing Technolology 23 (2004) 419–428.

[23] L.M. Owsley, L.E. Atlas, G.D. Bernard, Self-organizing feature maps and hidden Markov models for machine-tool monitoring, IEEE Transactions on Signals Processing 45 (1997) 2787–2798.

[24] B. Sick, On-line and indirect tool wear monitoring in turning with artificial neural network: a review of more than a decade of research, Mechanical System Signal Processing 16 (2002) 487–546.

[25] L Cabulea, M. Aldea, Making a decision when dealing with uncertain conditions, Acta Universitatis Apulensis Mathematics-Informatics 7 (2004) 85–92.

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

[27] Z. Yao, D. Mei, Z. Chen, On-line chatter detection and identification based on wavelet and support vector machine, Journal of Materials Processing Technology 210 (5) (2010) 713–719.

[28] S. Tangjitsitcharoen, In-process monitoring and detection of chip formation and chatter for CNC turning, Journal of Materials Processing Technology 209 (10) (2009) 4682–4688.

[29] X. Li, A brief review: acoustic emission method for tool wear monitoring during turning, International Journal of Machine Tools and Manufacture 42 (2002) 157–165.