An Optimization Method for Training Generalized Hidden Markov Model based on Generalized Jensen Inequality

Y. M. Hu[1], F. Y. Xie[1,2], B. Wu[1], Y. Cheng[1], G. F. Jia[1], Y. Wang[3] and M. Y. Li[1]
[1]*State Key Laboratory for Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, P.R. China*
[2]*School of Mechanical and Electronical Engineering, East China Jiaotong University, Nanchang 330013, P.R. China*
[3]*Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, U.S.A.*

Keywords: Generalized Hidden Markov Model, Generalized Jensen Inequality, Generalized Baum-Welch Algorithm.

Abstract: Recently a generalized hidden Markov model (GHMM) was proposed for solving the problems of aleatory uncertainty and epistemic uncertainty in engineering application. In GHMM, the aleraory uncertainty is derived by the probability measure while epistemic uncertainty is modelled by the generalized interval. Given any finite observation sequence as training data, the problem of training GHMM is often encountered. In this paper, an optimization method for training GHMM, as a generalization of Baum-Welch algorithm, is proposed. The generalized convex and concave functions based on the generalized interval are proposed for inferring the generalized Jensen inequality. With generalized Baum-Welch's auxiliary function and generalized Jensen inequality, similar to the multiple observations training, the GHMM parameters are estimated and updated by the lower and the bound observation sequences. A set of training equations and re-estimated formulas have been derived by optimizing the objective function. Similar to multiple observations (expectation maximization) EM algorithm, this method guarantees the local maximum of the lower and the upper bound and hence the convergence of the GHMM training process.

1 INTRODUCTION

Jensen inequality, named Johan Jensen in 1906, relates the value of a convex function of an integral to the integral of the convex function (Jensen, 1906). As an important mathematical tool it has been widely used, such as probability density function, statistical physics, information theory, and optimization theory. However it does not differentiate two kinds of uncertainties, namely, aleatory uncertainty is inherent randomness, whereas epistemic uncertainty is due to lack of knowledge. Epistemic uncertainty is significant and cannot be ignored. The generalized interval provides a valid method for solving epistemic uncertainty. Compared to the classical interval, generalized interval based on the Kaucher arithmetic (Kaucher, 1980) is better in algebraic properties so that the calculus can be simplified (Wang, 2011).

As a generalization of hidden Markov model (HMM) (Rabiner, 1989), the (Generalized hidden Markov model) GHMM, which is based on the generalized interval probability, are stochastic models in capable of statistical learning and classification (Wang, 2011). The optimization of GHMM parameters is a crucial problem for the application of GHMMs, since it can create the best models for real phenomena. Similar to HMM, a generalized Baum-Welch algorithm can be adapted such that the result is the local maximum (Baum et al., 1970).

In this paper, an optimization method, which is for training GHMM based on the generalized jensen inequality and the generalization Baum-Welch algorithm, is proposed. The classical convex functions, the classical concave functions, and the Jensen inequality are respectively developed by the use of generalized intervals. The parameters of GHMM are estimated and updated through generalized Baum-Welch algorithm. A generalized Baum-Welch's auxiliary function is built up and the generalized Jensen inequality is used. Similar to the multiple observations training in HMM (Li et al., 2000), the GHMM parameters are estimated and updated by given the lower and the bound observation sequences. A set of training equations

M. Hu Y., Y. Xie F., Wu B., Cheng Y., F. Jia G., Wang Y. and Y. Li M..
An Optimization Method for Training Generalized Hidden Markov Model based on Generalized Jensen Inequality.
DOI: 10.5220/0004118202680274
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2012), pages 268-274
ISBN: 978-989-8565-21-1

have been derived by optimizing the objective function. A set of GHMM re-estimated formulas has been derived by the unique maximum of the objective function. The proposed optimization method takes advantage of the good algebraic property in the generalized interval. This provides an efficient approach to train the GHMM.

2 BACKGROUND

2.1 Generalized Interval

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 $\overline{x}$ (Popova, 2000, Gardenes, 2001). Let $\mathbf{x} := [\underline{x}, \overline{x}]$, $\underline{x} \geq 0, \overline{x} \geq 0$ $(\underline{x}, \overline{x} \in \mathbb{R})$ and $\mathbf{y} := [\underline{y}, \overline{y}]$, $\underline{y} \geq 0, \overline{y} \geq 0$ ($\underline{y}, \overline{y} \in \mathbb{R}$) be interval variables, and let generalized interval function be $\mathbf{f}(t) = [\underline{f}(t), \overline{f}(t)]$ (Markov, 1979), where t is a real variable, then, the arithmetic operations of generalized intervals based on the Kaucher arithmetic are follows.

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]. \quad (1)$$

$$\mathbf{x} \text{ - } dual\mathbf{y} = [\underline{x} - \underline{y}, \overline{x} - \overline{y}]. \quad (2)$$

$$\mathbf{x} \times \mathbf{y} = [\underline{x} \times \underline{y}, \overline{x} \times \overline{y}]. \quad (3)$$

$$\mathbf{x} \div dual\mathbf{y} = [\underline{x} \div \underline{y}, \overline{x} \div \overline{y}], \underline{y} \neq 0, \overline{y} \neq 0. \quad (4)$$

$$\log\mathbf{x} = [\log\underline{x}, \log\overline{x}], \underline{x} \neq 0, \overline{x} \neq 0. \quad (5)$$

$$\int \mathbf{f}(t)\mathrm{d}t = [\int \underline{f}(t)\mathrm{d}t, \int \overline{f}(t)\mathrm{d}t]. \quad (6)$$

$$d\,\mathbf{f}(t)/dt = \left[d\,\underline{f}(t)/dt, d\,\overline{f}(t)/dt \right]. \quad (7)$$

$$\frac{\partial}{\partial x}\mathbf{f}(t) = \lim_{\Delta x \to 0} \left(\mathbf{f}(t + \Delta t) - dual\mathbf{f}(t) \right) / \Delta t \quad (8)$$

The less than or equal to partial order relationship between two generalized intervals is defined as

$$[\underline{x}, \overline{x}] \geq [\underline{y}, \overline{y}] \Leftrightarrow \underline{x} \geq \underline{y} \wedge \overline{x} \geq \overline{y}. \quad (9)$$

2.2 Generalized Hidden Markov Model

The GHMM is a generalization of HMM in the context of generalized interval probability theory. The generalized interval probability is based on the generalized interval with a form of probability. In GHMM, all probability parameters of HMM are replaced by the generalized interval probabilities. The boldface symbols have generalized interval values.

A GHMM is characterized as follows. The values of hidden states are in the form of $S = \{S_1, S_2, \cdots, S_N\}$, where N is the total number of possible hidden states. The hidden state variable at time t is $\boldsymbol{q}_t$, where $\boldsymbol{q}_t := [\underline{q}_t, \overline{q}_t]$. The M possible distinct observation symbols per state are $V = \{v_1, v_2, \cdots, v_M\}$. The generalized observation sequence is in the form of $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots, \boldsymbol{o}_T)$ where $\boldsymbol{o}_t$ is the observation value at time t. Note that the observations have the values of generalized intervals. Equivalently the lower bound sequence $\underline{O} = (\underline{o}_1, \underline{o}_2, \cdots \underline{o}_T)$ and upper bound sequence can be viewed separately, where, the value of $\underline{o}_t$ and $\overline{o}_t$ (t=1,…,T) can be any of $\{v_1, v_2, \cdots, v_M\}$.

Let $q_t \in \mathrm{pro}[\underline{q}_t, \overline{q}_t]$ and $o_t \in \mathrm{pro}[\underline{o}_t, \overline{o}_t]$ be real-valued random variables. $\boldsymbol{A} = (\boldsymbol{a}_{ij})_{N \times N}$ is the state transition interval probability matrix, $\boldsymbol{a}_{ij} = \mathbf{p}(q_{t+1} = S_j \mid q_t = S_i)$, $(1 \leq i, j \leq N)$ 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. $\boldsymbol{b}_j(k) = \mathbf{p}(o_t = v_k \mid q_t = S_j)$, $(1 \leq j \leq N, 1 \leq k \leq M)$ is the interval probability of observations in state S_j at time t. $\boldsymbol{\pi} = (\boldsymbol{\pi}_i)_{1 \times N}$ is the initial state interval probability distribution, $\boldsymbol{\pi}_i = \mathbf{p}(q_1 = S_i), (1 \leq i \leq N)$. The compact GHMM is denoted as $\boldsymbol{\lambda} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$.

Similar to classical HMM, the GHMM is usually used to solve three basic problems 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 best models for real phenomena. The proposed generalized Baum-Welch algorithm, as a generalization of Baum-Welch algorithm in the context of the generalized interval probability, provides an efficient approach to train GHMM. The generalized Baum-Welch algorithm is based on the generalized Jensen inequality which is introduced in the following section.

3 GENERALIZED JENSEN INEQUALITY

3.1 Generalized Convex Function

Following the definition of interval function by Markov, a generalized convex function is defined as following (Markov, 1979)

A generalized interval function $\mathbf{f}(t)$ is in the domain of the generalized interval $\mathbb{L}$, where t is a real variable, $\forall t_1, t_2 \in \mathbb{L}$ if

$$\mathbf{f}(\mathbf{r}_1 t_1 + \mathbf{r}_2 t_2) \le \mathbf{r}_1 \mathbf{f}(t_1) + \mathbf{r}_2 \mathbf{f}(t_2) \tag{10}$$

Where $\mathbf{r}_1 := [\underline{r}_1, \overline{r}_1]$, $\mathbf{r}_2 := [\underline{r}_2, \overline{r}_2]$, $\underline{r}_1 \ge 0, \overline{r}_1 \ge 0$, $\underline{r}_2 \ge 0, \overline{r}_2 \ge 0$, and $\mathbf{r}_1 + \mathbf{r}_2 = [1,1]$. $\mathbf{f}(t)$ is named as generalized convex function in the generalized interval $\mathbb{L}$.

Under the same assumption condition as generalized convex function, if

$$\mathbf{f}(\mathbf{r}_1 t_1 + \mathbf{r}_2 t_2) \ge \mathbf{r}_1 \mathbf{f}(t_1) + \mathbf{r}_2 \mathbf{f}(t_2) \tag{11}$$

$\mathbf{f}(t)$ is named as generalized concave function in the generalized interval $\mathbb{L}$, especially, when $\mathbf{f}(t) = \log t$ (t>0), then, $\dfrac{d^2 \mathbf{f}(t)}{dt^2} = -t^2 < 0$, $\mathbf{f}(t)$ is named as a strict generalized concavity of *log* function.

3.2 Generalized Jensen Inequality

For a generalized convex function $\mathbf{f}(t)$, numbers $t_1, t_2, ..., t_n$ in its domain, 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

$$\mathbf{f}\left(\sum_{i=1}^{n} \mathbf{r}_i t_i\right) \le \sum_{i=1}^{n} \mathbf{r}_i \mathbf{f}(t_i) \tag{12}$$

A mathematical induction is adopted to prove formula (12).

Proof. When n=2, formula (12) is a defined form of generalized convex function, thus it is correct.

If n=k, $\mathbf{f}\left(\sum_{i=1}^{k} \mathbf{r}_i t_i\right) \le \sum_{i=1}^{k} \mathbf{r}_i \mathbf{f}(t_i)$ now we need to prove that formula (12) is also correct when n=k+1

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

Thus, we can obtain that formula (12) is correct for all $i\,(i = 1, 2, ..., n)$.

It is a simple corollary that the opposite is true for generalized concave function transformations, the generalized interval formula is

$$\mathbf{f}\left(\sum_{i=1}^{n} \mathbf{r}_i t_i\right) \ge \sum_{i=1}^{n} \mathbf{r}_i \mathbf{f}(t_i) \tag{13}$$

4 OPTIMIZATION METHOD IN TRAINING PROCESS OF A GHMM

The training problem of GHMM is that given the observation sequence $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, ..., \boldsymbol{o}_T)$, we adjust the model parameters $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$ to maximize the lower and the upper bound of $\mathbf{p}(\boldsymbol{O} \mid \boldsymbol{\lambda})$. Similar to the training of HMM, we can choose $\boldsymbol{\lambda} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$ so that the lower and the upper bound of $\mathbf{p}(\boldsymbol{O} \mid \boldsymbol{\lambda})$ are both locally maximized by using an iterative procedure of the generalized Baum-Welch algorithm. The optimization method is based on the following two lemmas similar to inference of HMM re-estimation formulas (Levinson et al., 1983).

Lemma 1: Let $\boldsymbol{u}_i =: [\underline{u}_i, \overline{u}_i]$, $\underline{u}_i > 0, \overline{u}_i > 0, i = 1, \dots, S$ be positive interval real numbers, and let $\boldsymbol{v}_i =: [\underline{v}_i, \overline{v}_i]$, $\underline{v}_i \ge 0, \overline{v}_i \ge 0, i = 1, \dots, S$ be nonnegative interval real numbers such that $\sum_i \boldsymbol{v}_i \ge [0,0]$. Then from the generalized concavity of the *log* function and the generalized Jensen inequality, it follows that

$$\begin{aligned}
\log \frac{\sum_i \boldsymbol{v}_i}{dual \sum_i \boldsymbol{u}_i} &\ge \sum_i \left(\frac{\boldsymbol{u}_i}{dual \sum_k \boldsymbol{u}_k} . \log \frac{\boldsymbol{v}_i}{dual \boldsymbol{u}_i} \right) \\
&= \frac{1}{dual \sum_k \boldsymbol{u}_k} . \left[\sum_i (\boldsymbol{u}_i \log \boldsymbol{v}_i - dual \boldsymbol{u}_i \log \boldsymbol{u}_i) \right]
\end{aligned} \tag{14}$$

Here every summation is from 1 to S.

Lemma 2: If $\boldsymbol{c}_i > 0, i = 1,\ldots,N$, then subject to the constraint $\sum_{i=1}^{N} t_i = 1$, the generalized interval function is

$$\mathbf{f}(t) = \sum_{i=1}^{N} \boldsymbol{c}_i \log t_i \tag{15}$$

We can obtain its unique global maximum when

$$t_i = \boldsymbol{c}_i \Big/ dual \sum_{i=1}^{N} \boldsymbol{c}_i \tag{16}$$

The proof of formulas (16) is similar to Lagrange method (Levinson et al., 1983)

$$\frac{\partial}{\partial x_i}\left[\mathbf{f}(t) - dual\,\boldsymbol{\mu} \sum_i t_i\right] = \frac{\boldsymbol{c}_i}{dual x_i} - \boldsymbol{\mu} = 0 \tag{17}$$

Multiplying by t_i and summing over i, we can obtain $\boldsymbol{\mu} = \sum_i c_i$, so formula (16) can be derived.

4.1 Auxiliary Interval Function

In order to train the GHMM, the lower bound observation sequence $\underline{O} = (\underline{o}_1, \underline{o}_2, \ldots \underline{o}_T)$, for example, is used as the input of auxiliary interval function $\boldsymbol{Q}^l(\lambda, \lambda)$, which is defined as following.

Let $\boldsymbol{u}_i^l$ be the joint interval probability $\boldsymbol{u}_i^l := \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda)$ which is depended on model λ, and let $\boldsymbol{v}_i^l$ be the same joint interval probability $\boldsymbol{v}_i^l := \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda)$ which is depended on model λ, where $\boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_T)$, then

$$\sum_i \boldsymbol{u}_i^l = \mathbf{p}(\underline{O} \mid \lambda), \sum_i \boldsymbol{v}_i^l = \mathbf{p}(\underline{O} \mid \lambda) \tag{18}$$

Let auxiliary interval function $\boldsymbol{Q}^l(\lambda, \lambda)$ (Levinson et al., 1983) be

$$\boldsymbol{Q}^l(\lambda, \lambda) = \sum_{\boldsymbol{Q}} \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) \log \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) \tag{19}$$

Let formula (18) substitute into formula (14) of Lemma 1

$$\begin{aligned} &\log \frac{\mathbf{p}(\underline{O} \mid \lambda)}{dual\mathbf{p}(\underline{O} \mid \lambda)} \\ &\geq \frac{1}{dual\mathbf{p}(\underline{O} \mid \lambda)} . \left[\sum \left(\boldsymbol{Q}(\lambda, \lambda) - dual\boldsymbol{Q}(\lambda, \lambda)\right)\right] \end{aligned} \tag{20}$$

In formula (20), we can obtain $\mathbf{p}(\underline{O} \mid \lambda) \geq \mathbf{p}(\underline{O} \mid \lambda)$ if $\boldsymbol{Q}^l(\lambda, \lambda) \geq \boldsymbol{Q}^l(\lambda, \lambda)$, i.e., if we can find a model λ that makes the right-hand side of formula (20) positive, we can find a way to improve the model λ. Clearly, the largest guaranteed improvement by this method results for λ, which maximizes $\boldsymbol{Q}^l(\lambda, \lambda)$, and hence maximizes the lower and the upper bound of $\mathbf{p}(\underline{O} \mid \lambda)$.

4.2 Training of the Lower and the Upper Bound

Similar to Baum-Welch algorithm in HMM (Levinson et al., 1983), the training formulas can be inferred as following.

$$\begin{aligned} \log \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) &= \log\left(\mathbf{p}(\boldsymbol{Q} \mid \lambda) . \mathbf{p}(\underline{O} \mid \boldsymbol{Q}, \lambda)\right) \\ &= \log \overset{l}{\boldsymbol{\pi}}_{\boldsymbol{q}_1} + \sum_{t=1}^{T-1} \log \overset{l}{\boldsymbol{a}}_{\boldsymbol{q}_t \boldsymbol{q}_{t+1}} + \sum_{t=1}^{T} \log \overset{l}{\boldsymbol{b}}_{\boldsymbol{q}_t}(\underline{O}_t) \end{aligned} \tag{21}$$

Substituting this into formula (19), and re-grouping terms in the summations according to state transitions and observations, it can be seen that

$$\boldsymbol{Q}(\lambda, \lambda) = \sum_{i=1}^{N} \sum_{i=1}^{N} \boldsymbol{c}_{ij}^l \log \overset{l}{\boldsymbol{a}}_{ij} + \sum_{i=1}^{N} \sum_{i=1}^{N} \boldsymbol{d}_{jk}^l \log \overset{l}{\boldsymbol{b}}_{j(k)} + \sum_{t=1}^{N} \boldsymbol{e}_1^l \log \overset{l}{\boldsymbol{\pi}}_i \tag{22}$$

Where

$$\boldsymbol{c}_{ij}^l = \sum_{\boldsymbol{Q}} \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) . \sum_{t=1}^{T-1} \xi_t^l(i, j) = \mathbf{p}(\underline{O} \mid \lambda) . \sum_{t=1}^{T-1} \xi_t^l(i, j) \tag{23}$$

$$\boldsymbol{d}_{jk}^l = \sum_{\boldsymbol{Q}} \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) . \sum_{t=1, \underline{O}_t = v_k}^{T} \gamma_t^l(j) = \mathbf{p}(\underline{O} \mid \lambda) . \sum_{t=1, \underline{O}_t = v_k}^{T} \gamma_t^l(j) \tag{24}$$

$$\boldsymbol{e}_1^l = \sum_{\boldsymbol{Q}} \mathbf{p}(\underline{O}, \boldsymbol{Q} \mid \lambda) . \gamma_1^l(i) = \mathbf{p}(\underline{O} \mid \lambda) . \gamma_1^l(i) \tag{25}$$

Where $\xi_t^l(i, j)$ is the lower interval probability of being in state S_i at time t and in state S_j at time $t+1$, $\gamma_t^l(i)$ is the lower interval probability of being in state S_i at time t, $\gamma_1^l(i)$ is the lower interval probability of being in state S_i at the beginning of the observation sequence.

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 formulas (16), $\boldsymbol{Q}^l(\lambda, \lambda)$ is maximized if only if

$$\boldsymbol{a}_{ij}^{l} = \frac{\boldsymbol{c}_{ij}^{l}}{dual\sum_{j}\boldsymbol{c}_{ij}^{l}} = \sum_{t=1}^{T-1}\xi_t^l(i,j) \Big/ dual\sum_{t=1}^{T-1}\gamma_t^l(i) \quad (25)$$

$$\boldsymbol{b}_{j(k)}^{l} = \boldsymbol{d}_{jk}^{l} \Big/ dual\sum_{k}\boldsymbol{d}_{jk}^{l} = \sum_{t=1, \underline{o}_t = v_k}^{T}\gamma_t^l(j) \Big/ dual\sum_{t=1}^{T}\gamma_t^l(j) \quad (26)$$

$$\boldsymbol{\pi}_i^{l} = \boldsymbol{e}_1^{l} \Big/ dual\sum_{i}\boldsymbol{e}_1^{l} = \gamma_1^l(i) \quad (27)$$

Where $\boldsymbol{a}_{ij}^{l}$, $\boldsymbol{b}_{j(k)}^{l}$ and $\boldsymbol{\pi}_i^{l}$ are respectively the lower state transition probability, the lower observation probability, and the lower initial state probability used in the form of the generalized interval.

These are regarded as the lower bound re-estimation formulas. The maximum value of $\boldsymbol{Q}^l(\boldsymbol{\lambda},\boldsymbol{\lambda})$ can be reached by the lower bound re-estimation formulas. And hence the maximum value of $\mathbf{p}(\underline{O}\,|\,\boldsymbol{\lambda})$ is also obtained.

By the same method, we can obtain the upper bound re-estimation formulas (28) ~ (30). The maximum value of $\boldsymbol{Q}^u(\boldsymbol{\lambda},\boldsymbol{\lambda})$ can be reached by the upper bound re-estimation formulas. And hence the maximum value of $\mathbf{p}(\overline{O}\,|\,\boldsymbol{\lambda})$ is also obtained.

$$\boldsymbol{a}_{ij}^{u} = \boldsymbol{c}_{ij}^{u} \Big/ dual\sum_{j}\boldsymbol{c}_{ij}^{u} = \sum_{t=1}^{T-1}\xi_t^u(i,j) \Big/ dual\sum_{t=1}^{T-1}\gamma_t^u(i) \quad (28)$$

$$\boldsymbol{b}_{j(k)}^{u} = \boldsymbol{d}_{jk}^{u} \Big/ dual\sum_{k}\boldsymbol{d}_{jk}^{u} = \sum_{t=1, \overline{o}_t = v_k}^{T}\gamma_t^u(j) \Big/ dual\sum_{t=1}^{T}\gamma_t^u(j) \quad (29)$$

$$\boldsymbol{\pi}_i^{u} = \boldsymbol{e}_1^{u} \Big/ dual\sum_{i}\boldsymbol{e}_1^{u} = \gamma_1^u(i) \quad (30)$$

Where $\boldsymbol{a}_{ij}^{u}$, $\boldsymbol{b}_{j(k)}^{u}$ and $\boldsymbol{\pi}_i^{u}$ 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.

4.3 Training of the GHMM

According to the concept of multiple observation sequences (Li et al,. 2000), $\underline{O} = (\underline{o}_1, \underline{o}_2, \cdots \underline{o}_T)$ and $\overline{O} = (\overline{o}_1, \overline{o}_2, \cdots, \overline{o}_T)$ are regarded as two independence observation sequences, a group of GHMM re-estimation formulas can be defined according to the EM algorithm as follows.

$$\boldsymbol{a}_{ij} = \sum_{t=1}^{T-1}\xi_t^l(i,j) + \sum_{t=1}^{T-1}\xi_t^u(i,j) \Big/ dual\left(\sum_{t=1}^{T-1}\gamma_t^l(i) + \sum_{t=1}^{T-1}\gamma_t^u(i)\right) \quad (31)$$

$$\boldsymbol{b}_{j(k)} = \sum_{t=1, \underline{o}_t = v_k}^{T}\gamma_t^l(j) + \sum_{t=1, \overline{o}_t = v_k}^{T}\gamma_t^u(j) \Big/ dual\left(\sum_{t=1}^{T}\gamma_t^l(j) + \sum_{t=1}^{T}\gamma_t^u(j)\right) \quad (32)$$

$$\boldsymbol{\pi}_i = \frac{1}{2}\left(\gamma_1^l(i) + \gamma_1^u(i)\right) \quad (33)$$

Where $\boldsymbol{a}_{ij}$, $\boldsymbol{b}_{j(k)}$ and $\boldsymbol{\pi}_i$ are the state transition interval probability, the observation interval probability, and the initial state interval probability, respectively.

The training model parameters $\boldsymbol{\lambda} = \{\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}\}$ can be obtained by re-estimation formulas (31) ~ (33). With $\underline{O} = (\underline{o}_1, \underline{o}_2, \cdots \underline{o}_T)$ and $\overline{O} = (\overline{o}_1, \overline{o}_2, \cdots, \overline{o}_T)$ regarded as two independence observation sequences, we can define

$$\mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda}) := \mathbf{p}(\underline{O}\,|\,\boldsymbol{\lambda}) \times \mathbf{p}(\overline{O}\,|\,\boldsymbol{\lambda}) \quad (34)$$

$$\mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda}) := \mathbf{p}(\underline{O}\,|\,\boldsymbol{\lambda}) \times \mathbf{p}(\overline{O}\,|\,\boldsymbol{\lambda}) \quad (35)$$

According to the inference of the lower and the upper bounds re-estimation formulas, $\mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda}) \geq \mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda})$ can be also obtained for the values of interval probabilities are between 0 and 1.

The re-estimation formulas are derived by Lagrange interval method which guarantees the convergence of the GHMM training process. The local maxima can be derived by the iterative procedure of GHMM re-estimation formulas.

The iterative procedure for finding the optimal model parameters is:

- Choose an initial model $\boldsymbol{\lambda} = \{\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}\}$.
- Choose the generalized observation sequence $\underline{O} = (\underline{o}_1, \underline{o}_2, \cdots \underline{o}_T)$, $\overline{O} = (\overline{o}_1, \overline{o}_2, \cdots, \overline{o}_T)$.
- Obtain the training model $\boldsymbol{\lambda} = \{\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}\}$ which is determined by formulas (31 ~ 33).
- If local or global optimal of both $\underline{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda})$ and $\overline{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda})$ $\left(\mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda}) := \left[\underline{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda}), \overline{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda})\right]\right)$ are reached, then stop; otherwise, go back to step 3) and use $\boldsymbol{\lambda}$ in place of $\boldsymbol{\lambda}$.

The final result $\mathbf{p}(\boldsymbol{O}\,|\,\boldsymbol{\lambda})$ is so called a maximum likelihood estimation of GHMM.

4.4 Case Studies

In order to demonstrate the convergence of the proposed the optimization method for training GHMM, a training model of tool wear is studied. The experimental bench is illustrated in Figure 1.

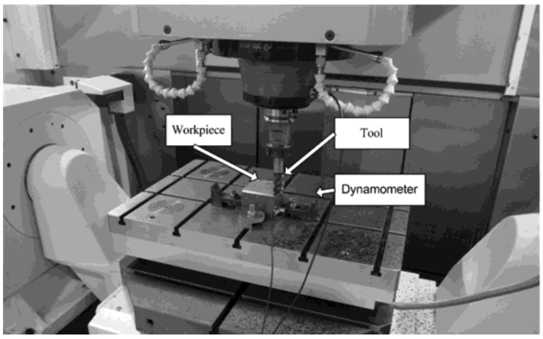

Figure 1: Experimental bench for cutting processing.

The cutting tests were conducted on Mikron UCP800 Duro, which is a five axis machining centre. The thrust force was measured by a Kistler 9253823 dynamometer. The resulting signals are converted into output voltages and then these voltage signals are amplified by Kistler multichannel charge amplifier 5070. Force signals were simultaneously recorded by NI PXIe-1802 data recorder. A 300M steel work piece material was adopted. The spindle speed was kept constant at 1000rpm and the feed rate was 400mm/min. The cutting depth 2mm and wide 2mm was adopted. A real time cutting signal with a dull tool processing is shown as Figure 2.

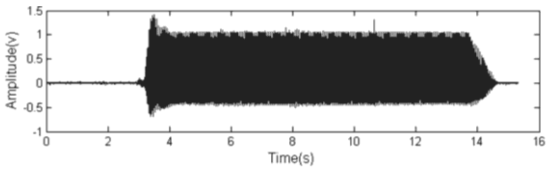

Figure 2: The cutting processing in time domain.

Time domain signals are considered as general error ±5%, and then the wavelet packet decomposition is used. The root mean square values of the wavelet coefficients at different scales were taken as the feature observations vector. The training procedure for finding the optimal model is carried out. The convergence curve of *log* likelihood is shown as Figure 3, and hence the convergence of the GHMM training process can be obtained.

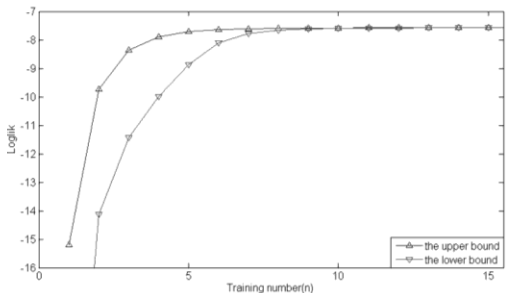

Figure 3: The training convergence curve.

5 CONCLUSIONS

Two kinds of uncertainties in engineering application can be encountered. The aleraory uncertainty is derived by the probability measure while the epistemic uncertainty is modelled by the generalized interval in GHMM. In this paper, the generalized convex and concave functions based on the generalized interval are proposed for inferring the generalized Jensen inequality. An optimization method for training GHMM, as a generalization of Baum-Welch algorithm, is proposed. The observation sequence is viewed separately as the lower and the bound observation sequences. The generalized Baum-Welch's auxiliary function and generalized Jensen inequality are used. Similar to HMM training, a set of training equations has been derived by optimizing the objective function. The lower and upper bound re-estimated formulas have been derived by unique maximum of the objective function. With a multiple observation concept, a group of GHMM re-estimation formulas has been derived. According to multiple observations EM algorithm, this method guarantees the local maximum of the lower and upper bound and hence the convergence of the GHMM training process.

ACKNOWLEDGEMENTS

This research is supported by National Key Basic Research Program of China (973 Program, Grant No.2011CB706803), Natural Science Foundation of China (Grant No. 51175208), and Natural Science Foundation of China (Grant No. 51075161).

REFERENCES

Jensen, J., L. W. V., 1906. Sur les functions convexes et les inegalites entre les valeurs moyennes, *Acta Math*, 30: 175–193.

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

Kaucher, E., 1980. Interval analysis in the extended interval space IR. Computing Supplement, 2:33–49.

Wang, Y., 2011. Multiscale uncertainty quantification based on a generalized hidden Markov model. *ASME Journal of Mechanical Design*, 3: 1–10.

Baum, L. E., Petrie, T, Soules, G., Weiss, Norman., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist*, 41(1): 164–171.

Li, X. L., Parizeau, M., Plamondon, R., 2000. Training hidden Markov models with multiple observations – A

vombinatorial method. *IEEE Transactions on PAMI*, (22)4: 371–377.
Popova, E. D., 2000. All about generalized interval distributive relations. I. *Complete Proof of the Relations* Sofia.
Gardenes, E., Sainz, M. A., Jorba, L. Calm, R., Estela, R., Mielgo, H., and Trepat, A., 2001. Modal intervals. *Reliable Computing*, 2: 77–111.
Markov, S., 1979. Calculus for interval functions of a real variable *Computing*, 22(4): 325–337.
Levinson, S., Rabiner, R. and Sondhi. M., 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell Systems Technical Journal*, 62: 1035-1074.