Signal Processing 94 (2014) 319–329

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

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

Article history:
Received 20 November 2012
Received in revised form
3 May 2013
Accepted 12 June 2013
Available online 16 July 2013

Keywords:
Generalized hidden Markov model
Generalized Jensen inequality
Generalized Baum–Welch algorithm
Generalized interval probability
State recognition

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).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.06.009

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 $\mathcal{A}$ of random events over Ω, the generalized interval probability $\mathbf{p} \in \mathbb{KR}$ is defined as $\mathbf{p}$: $\mathcal{A} \to [0,1] \times [0,1]$ which obeys the axioms of Kolmogorov: (1) $\mathbf{p}(\Omega) = [1, 1]$;(2) $[0, 0] \leq \mathbf{p}(E) \leq [1, 1]$ $(\forall E \in \mathcal{A})$; 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 $S = \{S_1, S_2, \ldots, S_N\}$, 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 are $V = \{v_1, v_2, \ldots, v_M\}$. The generalized observation sequence is in the form of $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T)$ where $\boldsymbol{o}_t$ 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. $\underline{O} = (\underline{o}_1, \underline{o}_2, \ldots, \underline{o}_T)$ denotes the lower bound of the observation sequence, and $\overline{O} = (\overline{o}_1, \overline{o}_2, \ldots, \overline{o}_T)$ denotes the upper bound, where the value of $\underline{o}_t$ and $\overline{o}_t$ $(t=1,\ldots,T)$ can be any of $\{v_1, v_2, \ldots, v_M\}$.

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 $[\underline{o}_t, \overline{o}_t]$. $\boldsymbol{A} = (\boldsymbol{a}_{ij})_{N \times N}$ is the