Expert Systems with Applications 41 (2014) 744–751

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Increasing mapping based hidden Markov model for dynamic process monitoring and diagnosis

Zefang Li [a,b], Huajing Fang [a,*], Lisha Xia [a]

[a] Department of Control Science and Engineering, Huazhong University of Science and Technology, 1037 Luo-yu Road, Wuhan 430074, PR China
[b] Institute of Computer Technology and Software Engineering, Wuhan Polytechnic, 463 Guan-shan Road, Wuhan 430047, PR China

ARTICLE INFO

Keywords:
Increasing mapping
Hidden Markov model
Process monitoring
Fault diagnosis
Independent component analysis

ABSTRACT

Hidden Markov models (HMMs) perform parameter estimation based on the forward–backward (FB) procedure and the Baum–Welch (BW) algorithm. The two algorithms together may increase the computational complexity and the difficulty to understand the algorithm structure of HMMs clearly. In this study, an increasing mapping based hidden Markov model (IMHMM) is proposed. Between the observation sequence and possible state sequence an increasing mapping is established. The re-estimation formulas for the model parameters are derived straightforwardly based on these mappings instead of FB variables. The IMHMM has simpler algorithm structure and lower storage requirement than the HMM. Based on IMHMM, an expandable process monitoring and fault diagnosis framework for large-scale dynamical process is developed. To characterize the dynamic process, a novel index considering serial correlation is used to evaluate process state. The presented methodology is carried out in Tennessee Eastman process (TEP). The results show improvement over HMM in terms of memory complexity and training time of the model. Also, the power of IMHMM can be observed compared with principal component analysis (PCA) based methods.

1. Introduction

Hidden Markov models (HMMs) have become an effective statistical learning tool in various applications such as speech recognition (Rabiner, 1989), signal processing (Fan & Xia, 2001), handwriting recognition (Nel, du Preez, & Herbst, 2005), computational biology (Krogh, Brown, Mian, Sjolander, & Haussler, 1994). The HMM is capable of modeling dynamical dependencies and correlations between complex real world phenomena. The forward–backward (FB) algorithm which is based on dynamic programming methods is the foundation for estimating HMM parameters (Baum & Egon, 1967; Baum & Sell, 1968). In fact, the FB algorithm was presented earlier by Chang and Hancock (1966), although it is often attributed to Baum and his colleagues. The FB variables are introduced to compute the smoothed state densities, in terms of which HMM parameters can be updated according to the Baum–Welch (BW) algorithm (Baum, 1972; Baum, Petrie, Soules, & Weiss, 1970).

In spite of its complexity and numerical instability, the FB algorithm has been found useful in many areas including bioinformatics, computer security and signal processing, speech recognition (Austin, Schwartz, & Placeway, 1991). However, only when the FB algorithm is incorporated with the BW algorithm the HMM parameters can be re-estimated. This may increase the complexity of algorithm structure and the amount of computation and storage while training an HMM. In particular, the re-estimation procedure may be computationally expensive for estimation from long observation sequences (Khreich, Granger, Miri, & Sabourin, 2010; Meyer & Durbin, 2004; Warrender, Forrest, & Pearlmutter, 1999).

In order to reduce the complexity of HMMs, some efforts which seek to improve the FB algorithm have been reported in literature. The checkpointing algorithm saves only some reference columns or checkpoints rather than all the filtered state densities (Grice, Hughey, & Speck, 1997; Khreich et al., 2010; Tarnas & Hughey, 1998). The Forward-only algorithm directly propagates all smoothed information in forward-only manner (Churbanov & Winters-Hilt, 2008; Miklos & Meyer, 2005). Yet these two methods both reduce the memory complexity at the expense of increasing the computational complexity. The Forward Filtering Backward Smoothing (FFBS) algorithm (Ott, 1967; Raviv, 1967) is an alternative to FB with the same memory complexity and slightly fewer computations than FB. A modification to the FFBS termed Efficient Forward Filtering Backward Smoothing is developed to make the memory complexity independent of the length of observation sequence without incurring computational overhead (Khreich et al., 2010). But it requires the inverse of the transposed transition matrix, and at every time instant the filtered state densities are divided by the emission probabilities that are usually very small, which may result in singularity and numerical instability. In addition,

* Corresponding author. Tel.: +86 18971062985.
E-mail addresses: 408729494@qq.com (Z. Li), hjfang@mail.hust.edu.cn (H. Fang).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.07.098

there are approximations to FB (Florez-Larrahondo, 2005; Hammer & Tjelmeland, 2011; Wang, Guan, & Zhang, 2004), whereas they result in different HMMs and their theoretical convergence properties are convinced only when the data sequence is infinite.

Most of these alternatives, however, can be viewed as variants of the FB algorithm in nature because they all propagate state information by the forward or backward pass. Their difference mainly lies in which type of state densities (filtered, smoothed, predictive or their combinations) they choose. In other words, the algorithm structure of HMMs based on BW and one of the variants of FB is still complex. In this paper, an increasing mapping based hidden Markov model (IMHMM) is first proposed and applied to perform dynamical process monitoring and fault diagnosis in large-scale process. In contrast with HMM and its extensions, such as hidden semi-Markov model (Dong & He, 2007; Yu & Kobayashi, 2006), autoregressive HMM (Rabiner, 1989), IMHHM does not involve any FB variables or their variants. The re-estimation formulas are derived through the joint probability of the observation sequence and increasing mapping which stems from combinatorial theory in mathematics (Mao, 1990). The memory complexity of IMHHM is reduced and the training time of the model is shortened in most cases due to its simpler algorithm structure inherently than that of HMM.

This paper is organized with the following fashion. Section 2 briefly introduces the HMM, the increasing mapping and independent component analysis (ICA). In Section 3, the novel increasing mapping based HMM is presented. Following that, we develop an IMHMM based process monitoring and diagnosis framework. Section 4 further validates the proposed system using data collected from TEP. Finally, conclusions are drawn in Section 5.

2. Theoretical background

2.1. Elements of HMM

The HMM is a finite-state discrete-time first-order Markov chain characterized by two related mechanisms, i.e., a homogeneous Markov chain with unobservable states and a set of emission probabilities depending on the current state. The fundamental elements of HMM are specified by the following (Rabiner, 1989):

(1) A set of N hidden states $S = \{s_i\}$, $i = 1,2,\ldots,N$.
(2) The state transition probability distribution $A = \{a_{ij}\}$ with

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$$

where q_t is the hidden state visited at time t.
(3) A set of observation symbols $V = \{v_1, v_2, \ldots, v_M\}$, where M is the number of distinct observations for each state.
(4) The observation probability distribution, one for each state, $B = \{b_i(k)\}$, where

$$b_i(k) = P(v_k \text{ at } t | q_t = s_i), \quad 1 \leqslant i \leqslant N, 1 \leqslant k \leqslant M.$$

(5) An initial state distribution $\pi = \{\pi_i\}$, where

$$\pi_i = P(q_1 = s_i), \quad 1 \leqslant i \leqslant N.$$

Hence, the complete parameter set of HMM is often represented by the compact notation: $\lambda = (\pi, A, B)$.

2.2. Increasing mapping

In order to give a precise definition of the increasing mapping, it is important to know what is meant in combinatorial theory by a relation. *A relation* on a set X is a subset R of the set $X \times X$ which is a set of ordered pairs of elements in X. Some special properties (Brualdi, 2002) for a relation R on X are as follows:

(1) R is reflexive, provided $(x,x) \in R$ for all x in X.
(2) R is antisymmetric, provided for x and y with $x \neq y$, if $(x,y) \in R$, then $(y,x) \notin R$.
(3) R is transitive, provided that if $(x,y) \in R$ and $(y,z) \in R$, then $(x,z) \in R$.

A partial order on X is a reflexive, antisymmetric and transitive relation R. A partial order R is *a total order*, provided for all x and y in X, either $(x,y) \in R$ or $(y,x) \in R$.

Definition 1. Let R_1, R_2 be a partial order on X and A, respectively. The mapping $f: X \to A$ is an increasing mapping, provided that if $(x,y) \in R_1$, then $(f(x),f(y)) \in R_2$ (Mao, 1990).

For convenience, we usually denote a partial order R by $\leqslant$, and $(x,y) \in R$ becomes $x \leqslant y$. If we denote R_1, R_2 by $\leqslant_1$ and $\leqslant_2$, respectively, f is an increasing mapping provided whenever $x \leqslant_1 y$ holds, $f(x) \leqslant_2 f(y)$ also holds.

2.3. Independent component analysis (ICA)

ICA is a recently developed method which can extract underlying factors or components from multivariate statistical data typically in the form of a large database of samples. The basic idea of ICA is to decompose observed data into liner combinations of statistically independent components (Comon, 1994). Hyvarinen (1999) presented a simple and efficient fixed-point algorithm for ICA, called Fast ICA, where it is assumed that d measured variables can be expressed as liner combinations of m ($\leqslant d$) unknown independent components (ICs). The relationship is given by

$$X = AY + E \tag{1}$$

where $X = [x(1),x(2),\ldots,x(n)] \in R^{d \times n}$ is the data matrix of n samples, $A = [a_1, a_2, \ldots, a_m] \in R^{d \times m}$ is the unknown mixing matrix, $Y = [y(1),y(2),\ldots,y(n)] \in R^{m \times n}$ is the independent component matrix, and $E \in R^{d \times n}$ is the residual matrix. The ICs and the measured variables have means of zero. Equivalently, the objective of ICA is to find a demixing matrix W that is expressed as

$$\hat{Y} = WX \tag{2}$$

where the rows of reconstructed matrix $\hat{Y}$ are as independent of each other as possible. In terms of the assumption that the rows of W with the largest sum of squared coefficient have the greatest impact on the variation of $\hat{Y}$, the data dimension can be reduced by selecting a few rows of W. Denote the new matrix constituted by the selected a rows of W as W_d (dominant part of W) and the remaining rows of W as W_e (excluded part of W), respectively. Similarly, split the columns of A into two matrixes A_d and A_e. The ICA statistics I^2, I_e^2 and the squared prediction error Q^2 also known as the SPE statistic, are defined as follows:

$$I^2 = \hat{S}_d^T \hat{S}_d, \quad \hat{S}_d = W_d x \tag{3}$$

$$I_e^2 = \hat{S}_e^T \hat{S}_e, \quad \hat{S}_e = W_e x \tag{4}$$

$$Q^2 = e^T e = (x - \hat{x})^T (x - \hat{x}), \quad \hat{x} = A_d W_d x \tag{5}$$

The utilization of I^2 and I_e^2 enables the whole space spanned by the original variables to be monitored based upon a new foundation, and the Q^2 statistic monitors the residual part of the process variation (Lee, Yoo, & Lee, 2004).