Computers and Chemical Engineering 60 (2014) 339–353

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Bayesian and Expectation Maximization methods for multivariate change point detection

Marziyeh Keshavarz, Biao Huang*

Department of Chemical and Material Engineering, University of Alberta, Edmonton, Alberta, Canada T6G-2G6

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form 20 August 2013
Accepted 16 September 2013
Available online 1 October 2013

Keywords:
Change point detection
Bayesian inference
Expectation Maximization

ABSTRACT

Process data are the most important information in all aspects of plant monitoring and control applications. These data, stemming from instruments, carry the necessary information that assists plant operations. One of the common problems of process instrument readings is their deviation from true values due to instrument bias or systematic error. Detection of change points in process data is the first step for a more insightful analysis of hidden factors affecting the process. In this paper, both Bayesian and Expectation and Maximization (EM) methods are considered for change point detection problem of multivariate data with both single and multiple changes. The performance of EM is compared with the Bayesian approach. Simulation results show superiority of EM in the case of improper selection of priors while the Bayesian approach has less computation demand. The proposed algorithms are evaluated through several examples, two from simulated random data and one from a CSTR problem. It is also verified through an experimental study of a hybrid tank system.

1. Introduction

Process data acquired from instruments are the basis for plant operation monitoring and control. Accuracy of these measurements is critical for normal and safe process operations. Consequently, timely detecting and repairing of faulty instruments are important for any process operation. Instrument bias, if occurs, remains until a corrective action such as recalibration is taken on the instrument. Detection of bias is especially important for reliable instrument applications. Instrument problems along with other factors such as a major process upset a change in equipment performance etc. may all introduce mean shift so that a bias is introduced in steady state operating data.

There are various techniques in literature to detect and estimate bias error in the instruments. An excellent review of single or multiple gross error detection methods along with their application can be found in Narasimhan and Jordache (2000). In all of these methods, process constraints such as material balance, energy balance, etc. are considered.

On the other hand, data-based approach to detection of the time instant at which bias is introduced has received great attention especially in applications such as hydrology, finance, economics and meteorology (Basseville & Nikiforov, 1993). In chemical engineering applications, there are a lot of sensors or instruments which may be subject to bias. Detection of time instants where these biases are introduced is important in instrument fault identification. Detection of instants where system operating mode changes can be another application of change point detection. These problems are often formulated as change point detection. At these change points, the mean of data shifts. As detection of these change points is performed, new mean and hence bias magnitude can also be determined.

Various methods have been developed in literature to tackle the problem of change point detection. A good review of these methods can be found in Basseville and Nikiforov (1993), Lu, Mausel, Brondizio, and Moran (2004) and Reeves, Chen, wang, Lund, and Lu (2007). Among these approaches, probabilistic frameworks, such as Bayesian inference, have been applied in various areas. These approaches are powerful in the sense that one can incorporate priori knowledge in estimation of unknown parameters. In Tamhane, Iordache, and Mah (1988a, 1988b), a Bayesian approach is used to detect gross errors based on process models. This method is applied sequentially over various time periods of data by updating the priors and posteriors at the end of each period. Computation of this method for medium to large problems is intensive despite the modification made by the authors of Tamhane, Iordache, and Mah (1988b). In Devanathan, Vardeman, and Rollins (2005), a Bayesian decision rule is developed to detect the change point in univariate data which needs selection of prior distributions for unknown parameters and then derivation of posterior probability

* Corresponding author. Tel.: +1 780 4929016.
E-mail address: biao.huang@ualberta.ca (B. Huang).

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.09.012

of shift point given the data. The multivariate version of Bayesian single change point detection can be found in Perreault, Parent, Bernier, Bobee, and Slivitzky (2000), Djafari and Feron (2007), Zambadouglas and Hawkins (2006) and Son and Kim (2005).

There are also various approaches developed to solve the problem of multiple change points detection based on methods such as hypothesis testing (Venter & Steel, 1996), maximum likelihood (Hawkins, 2001) and a clustering-based algorithm called product partition model (PPM) (Barry & Hartigan, 1993; Crowley, 1997). Among these methods, Bayesian approach has been most widely adopted.

In any Bayesian framework, having selected a prior for unknown variables and a certain distribution for likelihood, the problem becomes solving the posterior distribution by multiplying the prior and the likelihood. There are two approaches proposed in literature to solve the change point detection problem based on the Bayesian approach. One approach relies on finding the mode of posterior probability called Maximum a Posteriori (MAP) approach which is optimization-based. The other is to calculate means of various posterior probabilities, which leads to integration calculation. In general, these integrations are difficult to solve analytically. Markov Chain Monte Carlo (MCMC) is often used which draws samples from posterior distributions. The sampling from posterior distribution is performed using various techniques such as Metropolis-Hasting or Gibbs sampler (Cheon & Kim, 2010; Djafari & Feron, 2007; Loschi et al., 2008). In Cheon and Kim (2010), Stochastic Approximation Monte Carlo (SAMC) is applied to multiple change point detection problem and SAMC performance is compared with reversible jump Markov Chain Monte Carlo approach (RJMCMC).

On the other hand, Expectation Maximization (EM) can be regarded as an iterative way to find the maximum likelihood. It mainly consists of two steps: expectation step and maximization step. The algorithm iterates between these two steps until convergence. EM uses distribution of missing or hidden variables when it is not easy to directly maximize the likelihood of the observed data. Some researchers have already used EM to detect change points (Bansal, Du, & Hamedani, 2008; Yildirim, Singh, & Doucet, in press). In Yildirim et al. (in press), a Sequential Monte Carlo (SMC) online EM algorithm is proposed to estimate the change point. In Bansal et al. (2008) an EM method is presented to estimate the distribution of change point. Unlike those EM approaches, the EM algorithm presented in this paper does not require heavy and complex computation and it is relatively easy to implement.

In this paper, a closed form solution to the Bayesian formulation of single and multiple change point detection problem is first considered for multivariate data and MAP is used for the estimation of the parameters. Moreover, considering the sensitivity of the Bayesian approach to prior selection, EM is adopted to solve both single or multiple change points detection problem. By comparison, it is shown that EM is more powerful when priors are highly uncertain while the Bayesian approach has its advantage of less computation demand.

The main contributions of this paper are (1) extension of existing Bayesian method for change point detection by deriving a closed form analytical solution and (2) the derivation of EM algorithms for both single and multiple shift detection in multivariate data to overcome the limitations of Bayesian approach in the presence of improper priors.

The remainder of the paper is organized as follows. Section 2 gives an overview of multivariate Bayesian change point detection in the presence of single or multiple change points. Section 3 derives EM algorithm for solving the same problem. In Section 4, sensitivity of prior selection and proper initialization of EM are discussed. Finally, simulation results along with an experimental evaluation are provided to demonstrate the proposed algorithms.

2. Bayesian change point detection

2.1. Problem formulation for single change point

In this section, a multivariate Bayesian formulation of change point detection is provided where Maximum a Posteriori (MAP) is applied to infer the change point detection and mean. Throughout this paper, time instant for single change point is referred to as m and multiple time instants for multiple change points are represented by the vector $\boldsymbol{t}=[t_1, \ldots, t_N]$. Moreover, the covariance of data is assumed to be the same before and after the change points.

Consider that n observations from p variables form a $p \times n$ matrix as

$$D = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} & y_{1(m+1)} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2m} & y_{2(m+1)} & \cdots & y_{2n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pm} & y_{p(m+1)} & \cdots & y_{pn} \end{pmatrix} = (\boldsymbol{Y_1}, \boldsymbol{Y_2}, \ldots, \boldsymbol{Y_m}, \ldots, \boldsymbol{Y_n})$$

$Y_1, \ldots, Y_m, \ldots, Y_n$ are measurements of p variables from time instant 1 to n.

In next sections, we discuss the model for both single and multiple change points further.

2.2. Single change point detection

To model single change point, it is assumed that at the sampling instant m, a change occurs resulting in a shift in the mean vector. As a result, the whole data are split into two segments operating at two different means, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$, respectively with the same covariance matrix Σ. This general problem formulation framework is adopted throughout this work. Using Bayesian analysis, the objective is to find $P(m|D)$ which is the posterior probability of change point given the data. According to Bayes rule, this probability is equal to $P(m, D)/P(D)$. In the following, this posterior probability is derived in detail. It is assumed that observations are independent of each other and follow a normal distribution as

$$\begin{aligned} &\boldsymbol{Y_i} \sim N_p(\boldsymbol{\mu_1}, \Sigma), \quad i = 1, 2, \ldots, m \\ &\boldsymbol{Y}_i \sim \mathcal{N}_p(\boldsymbol{\mu_2}, \Sigma), \quad i = m+1, m+2, \ldots, n \end{aligned} \tag{1}$$

where $\mu_1 \neq \mu_2$. The normal distribution function can be expressed as

$$\mathcal{N}_p(\boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma^{-1}|^{1/2} exp\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\mu})\} \tag{2}$$

The shift time, m, can occur anywhere from 1 to $n-1$. The likelihood function of data, D, is therefore of the form

$$P(D|\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, m, \Sigma) = \prod_{i=1}^{m} P(\boldsymbol{Y}_i|\boldsymbol{\mu_1}, \Sigma) \prod_{i=m+1}^{n} P(\boldsymbol{Y}_i|\boldsymbol{\mu_2}, \Sigma) \tag{3}$$

where $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and m are to be determined. The priors for $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and m are taken as

$$\begin{aligned} P(\boldsymbol{\mu_1}|\boldsymbol{\mu_1^0}, \Sigma_{01}) &= \mathcal{N}_p(\boldsymbol{\mu_1^0}, \Sigma_{01}) \\ P(\boldsymbol{\mu_2}|\boldsymbol{\mu_2^0}, \Sigma_{02}) &= \mathcal{N}_p(\boldsymbol{\mu_1^0}, \Sigma_{02}) \\ P(m) &= \text{Uniform distribution} \end{aligned} \tag{4}$$