Computers and Chemical Engineering 69 (2014) 128–146

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Expectation Maximization method for multivariate change point detection in presence of unknown and changing covariance

Marziyeh Keshavarz, Biao Huang*

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

ARTICLE INFO

Article history:
Received 16 May 2013
Received in revised form 21 June 2014
Accepted 27 June 2014
Available online 16 July 2014

Keywords:
Change point detection
Bayesian inference
Expectation Maximization

ABSTRACT

Data analysis plays an important role in system modeling, monitoring and optimization. Among those data analysis techniques, change point detection has been widely applied in various areas including chemical process, climate monitoring, examination of gene expressions and quality control in the manufacturing industry, etc. In this paper, an Expectation Maximization (EM) algorithm is proposed to detect the time instants at which data properties are subject to change. The problem is solved in the presence of unknown and changing mean and covariance in process data. Performance of the proposed algorithm is evaluated through simulated and experimental study. The results demonstrate satisfactory detection of single and multiple changes using EM approach.

1. Introduction

Change point detection problem has received a great attention in various areas such as hydrology, economics, meteorology, pharmacology, and signal processing. This field was first introduced in quality control of manufacturing industry where control charts were used for quality assessment. In other applications such as climatology, analysis of instrumental data is of main interest as any level shifts occur due to instrumentation or relocation of sampling sites can lead to wrong interpretation of temporal trends (Caussinus and Mestre, 2004).

Various methods have been proposed in literature to solve the change point problems. Also, in practice for data segmentation applications, it is needed to detect multiple change points simultaneously. This becomes even more valuable if the two consecutive change points are close to each other. In the context of statistical approaches, these problems can be solved using either frequentist or Bayesian approach (Du, 2010). In frequentist approach, the focus is on likelihood. In other words, the analysis is performed on $P(D|H)$ which is probability of data given hypothesis. Here, the data are assumed to be random while the hypothesis is assumed to be fixed. In fact, in frequentist approach, one obtains the frequency with which one expects to observe the data, given the hypothesis. In Bayesian approach, however, the focus is on $P(H|D)$ which is the probability of a hypothesis given the data. The data are treated as fixed and the hypothesis as random.

Probabilistic methods based on likelihood have been used by many researchers (Hinkley, 1970; Hawkins, 2001). In Sen and Srivastava (1975) a binary segmentation method was applied assuming normally distributed data with a constant variance and different mean before and after the change. This method was further developed by Srivastava and Worsley (1986) for multivariate data. In Yao (1988), a method based on maximizing Schwarz criterion, i.e. a penalized likelihood, is proposed to select the partitions in data. In Hawkins (2001), the segmentation method was further generalized to multiple change points where a dynamic programming was employed. The extension of this solution to an unknown number of changes in the mean vector of multivariate data was derived in Caussinus and Mestre (2004).

Bayesian analysis of change point was first introduced in Chernoff and Zacks (1964). It was assumed that the shift in block mean to be a normally distributed increment with constant variance $N(\mu_i, \sigma 0)$ and hence a Markov model was introduced. A Bayesian method based on posterior probabilities of change points using binomial and normal distribution was developed in Smith (1975). The multivariate form of Bayesian single change point detection can be found in Perreault et al. (2000), Perreault et al. (2000), Djafari and Feron (2007), Zambadouglas and Hawkins (2006), Son and Kim (2005), and Karunamuni and Zhang (1996). In Karunamuni and Zhang (1996), an empirical Bayes stopping time is studied for detection of a change in distribution of data when prior is not completely known. Multiple change points

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

http://dx.doi.org/10.1016/j.compchemeng.2014.06.016
0098-1354/

detection is also investigated through methods such as hypothesis testing (Venter and Steel, 1996) and a clustering-based algorithm called product partition model (PPM).

Product partition model (PPM), is another method in change detection. In PPM technique, the prior probabilities for a random partition are determined. As a result, the posterior probability of the partition is of the same form. In essence, in this method, a large number of computations results in Markov sampling to determine the estimates of means, derived by conditioning on the partition and summing over all possible partitions (Barry and Hartigan, 1993; Crowley, 1997). The extension of PPM to find the probability of change in mean and variance of normal data using Gibbs sampling can be found in Loschi and Cruz (2005).

Generally, there are two approaches to solve the change point detection problem based on the Bayesian approach. One relies on finding the mode of posterior probability called maximum a posteriori (MAP) which is optimization-based. The other is to calculate means of various posterior probabilities leading to integration computation, which are difficult to solve analytically. As a result, 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 (Djafari and Feron, 2007; Cheon and Kim, 2010; Loschi et al., 2008). Markov sampling is a widely used method in calculation of Bayesian posterior probability. But according to Cheon and Kim (2010), in MCMC, due to many possible partitions, the model becomes complicated with multiple modes and hence traditional Monte Carlo methods may often get trapped in local energy minima.

In Cheon and Kim (2010), Stochastic Approximation Monte Carlo (SAMC) is applied to multiple change points detection problem and SAMC performance is compared with reversible jump Markov Chain Monte Carlo approach (RJMCMC). It was shown that in change point estimation, SAMC outperforms RJMCMC for complex Bayesian model selection problem.

In Lavielle (2006), the change points are determined by minimizing a penalized contrast function which measures how the model, derived based on change point sequence, fits the observed data.

On the other hand, Expectation Maximization (EM) can be viewed as an iterative approach to find the maximum likelihood (Gelman et al., 2004; MacLachlan and Krishnan, 1997). EM can also be employed to detect the changes. Some researchers have already used EM to detect change points (Yildirim et al., 2014; Bansal et al., 2008). In Yildirim et al. (2014), 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. These EM methods normally require complex calculation. In this paper, a new EM algorithm is proposed which does not require heavy computation and it is easy to implement as well. This proposed framework has the advantage of handling improper selection of hyperparameters compared with Bayesian approach.

In Keshavarz and Huang (2013, 2014), the change point problem was solved for mean shift detection with known and constant covariance in univariate and multivariate data respectively. In real process, however, the covariance of data is often unknown and varying. One way to deal with this problem is to estimate the covariance from the data. This approach, in the presence of several changes in the mean and covariance of data, may not work efficiently or may not work at all without knowing the change instances. An alternative way to handle this problem is through effective statistical inference approaches that can deal with unknown and changing covariance.

As mentioned in Keshavarz and Huang (2014), EM has a flexible framework through which several problems can be tackled. For instance, EM handles improper priors more effectively than the Bayesian method. This is one of the main advantages of using EM while it may be computationally heavier. In Keshavarz and Huang (2014), this advantage was demonstrated through several examples. In this paper, the focus is on derivation of EM solution for change point detection problem, in the presence of unknown covariance. Then, the solution is extended to a more general framework where both mean and covariance of data, as unknown parameters, can change simultaneously.

In the field of control performance monitoring, some methods are proposed for covariance change identification. These methods are based on generalized eigenvalue analysis and statistical inference (Yu and Qin, 2008, 2008). The difference between the proposed method in this paper for covariance change detection and those mentioned in Yu and Qin (2008, 2008) is that in proposed method, the data are not divided into monitoring and benchmark period. If the reference or benchmark data are not available, then the proposed approach in this paper can be a good alternative.

The main contribution of this paper is derivations of EM solution for both single and multiple shifts detection in multivariate data and in the presence of unknown and changing covariance. This method relies on steady state. As most of the methods in literature for bias detection, gross error detection, etc. are for steady state conditions. Through simulations and experimental study, it is shown that this method has satisfactory performance even in the presence of improper selection of priors.

The remainder of the paper is organized as follows. Section 2 gives a preliminary that is needed in the following sections. In Section 3, an introduction to EM algorithm is provided. EM derivations for single and multiple changes detection in the presence of unknown covariance is given in Section 4. In Section 5, the EM solution is derived for simultaneous mean and covariance change detection problem. Finally, simulation results along with an experimental evaluation are provided to validate the proposed algorithms.

2. Preliminary

Assume that in a sequence of independent and identically distributed (i.i.d) data, both mean and covariance are unknown and the objective is to estimate the mean and covariance of the data. A commonly used prior distribution for covariance matrix is Inverse Wishart (IW) distribution as

$$P(\Sigma|\nu_0, \Psi_0) = \frac{|\Psi_0|^{(\nu_0/2)}|\Sigma|^{-((\nu_0+p+1)/2)}\exp(-tr(\Psi_0\Sigma^{-1})/2)}{2^{(\nu_0 p/2)}Z(\nu_0, p)} \tag{1}$$

where

$$Z(n, p) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right) \tag{2}$$