Journal of Computational and Applied Mathematics 308 (2016) 447–463

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Detecting change-points for shifts in mean and variance using fuzzy classification maximum likelihood change-point algorithms

Kang-Ping Lu [a], Shao-Tung Chang [b,*]

[a] Department of Applied Statistics, National Taichung University of Science and Technology, Taichung, Taiwan

[b] Department of Mathematics, National Taiwan Normal University, Taipei, Taiwan

HIGHLIGHTS

- We propose a new algorithm, FCML-CP, to detect change-points in a statistical process.
- The mixture likelihood embedded fuzzy c-partition is utilized to estimate the change-points.
- The method outperforms statistical likelihood approaches.
- Various experiments show the effectiveness and practicability of the proposed method.
- Excellent performance in detecting small changes is helpful for root cause analyses.

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form 22 May 2016

Keywords:
Control chart
Change-point
Mixture model
Fuzzy clustering
Fuzzy classification maximum likelihood change-point algorithm

ABSTRACT

Knowing the time of changes, called change-point (CP), in a process is crucial for engineers to recognize the root cause fast and accurately. Since special causes may induce simultaneous changes in mean and variance, detecting changes in both at once is required. Many methodologies in quality control were developed for detecting changes in either mean or variance only, and process parameters were assumed known often. However, they are rarely known exactly and a small estimation error may lead to unfavorable CP estimates. Fuzzy partitioning is better suited to cases of vague boundaries between two segments which appear very often in reality. A new mechanism, called fuzzy classification maximum likelihood change-point (FCML-CP) algorithm, is proposed to detect shifts in mean and variance simultaneously. A CP framework is transferred into a mixture model and then a FCML-CP algorithm is created through fuzzy classification maximum likelihood procedures. The proposed FCML-CP can be applied to phase I and II processes without knowledge of in-control process parameters; it can estimate multiple CPs of process mean or/and variance simultaneously. The effectiveness and superiority of FCML-CP are shown by extensive experiments with numerical and real data sets. Specifically, the proposed FCML-CP is superior to the commonly used statistical mixture likelihood approach using expectation–maximization (EM) algorithm; it is much more time-saving especially. The remarkable performance of FCML-CP in detecting CPs for small changes is particularly important and helpful for engineers to recognize the special cause fast and correctly since an out-of-control signal resulted from small changes is usually delayed long.

* Corresponding author. Fax: +886 2 29332342.
E-mail address: schang4809@gmail.com (S.-T. Chang).

http://dx.doi.org/10.1016/j.cam.2016.06.006
0377-0427/

1. Introduction

Control charts are usually used in industries for monitoring processes and distinguishing the variation due to special causes. However, an out-of-control signal often delays long such that root cause analyses cannot be proceeded effectively. Knowing the time of a change, called a change-point (CP), is helpful for engineers to search for and identify the assignable causes fast and correctly. Hence, a good method for detecting CPs is essential for engineers to monitor processes efficiently. In quality engineering, processes control includes phase I and II analyses. In phase I, data are analyzed as a batch and multiple CPs are considered often. The work in phase I includes detecting CPs and estimating in-control process parameters, but in phase II, in-control process parameters are assumed known.

The CP problems in quality control were often solved by the maximum likelihood approach (see e.g., [1–10]). Recently, the artificial neural networks methodology was applied for CP detection (see [11–14]). Besides, a CP model framework for CP detection was considered by Hawkins et al. [15–17] and it was extended by Ross et al. [18,19] in various ways. A comprehensive literature review for CP estimation can be seen in [20]. Many CP methods in quality control were developed for detecting CPs for either mean or variance only, but few were designed for detecting changes in both at once. However, simultaneous changes in mean and variance may occur in reality. For instance, in a screen printing process, the thickness of the solder paste printed onto circuit boards may change in mean and variance simultaneously when a squeegee is not leveled appropriately (cf. [21]). Furthermore, the control limits in mean charts depend on the variance, hence, an out-of-control signal in a mean chart probably results from a change in variance but not in mean. In such a situation, a wrong conclusion may be drawn by a CP method for detecting changes in mean only. Thus, a method for identifying the time of changes in mean and variance simultaneously is necessary.

On the other hand, locating CPs in a process is akin to classifying data into clusters of similar individuals, hence, clustering techniques have also been applied to CP detection in the literature. Specifically, the mixture model is powerful and commonly used in cluster analysis, therefore, the mixture likelihood approach using the expectation and maximization (EM) algorithm has been employed to detect CPs in several studies. For example, Ng [22] employed an EM algorithm to efficiently locate the CP in a degradation process; Yildirima et al. [23] proposed an online EM algorithm for CP models; Keshavarz and Huang [24,25] applied EM algorithms for multivariate CP detection and they showed the superiority of EM over the Bayesian approach in cases of improper priors used. Yang [26] proposed a robust mean CP method in use of EM algorithm for regression analyses. Although the EM algorithm is applicable for CP detection, however, its slow rate of convergence restricts its use in cases of large data sets or multiple CPs. Furthermore, the EM algorithm may ended at a local maximum due to inappropriate initial values used.

As to the CP methods for simultaneously detecting changes in mean and variance in quality control, Hawkins and Zamba [27] introduced a generalized likelihood ratio test for detecting a CP in mean or variance for normal processes. Park and Park [28] presented a maximum likelihood estimator (MLE) of a CP for identifying a single shift in mean and/or variance. Park and Park's method is devoted to phase II analyses in which the true values of process parameters are assumed known. Unfortunately, these parameters are rarely known; they are usually estimated through phase I study instead. Any error associated with these estimates may lead to substantial increases in false alarm probabilities and unsatisfactory CP estimates. Moreover, Chang and Lu [29] made use of EM algorithms to create an EMCP algorithm to detect shifts in mean and variance simultaneously. They showed the mixture likelihood approach outperforms the traditional maximum likelihood method in terms of the accuracy and precisions of estimation; however, their proposed EMCP is extremely time consuming in detecting multiple CPs due to the slow rate of convergence. Motivated by the findings in [29] that the mixture likelihood approach for CP detection is superior but very time-consuming in use of EM algorithms, thus, we consider utilizing the fuzzy classification likelihood approach for CP detection to alleviate the problem of time consumption.

On the other hand, most clustering techniques used in CP detection adopted hard partitioning which restricts each data point belonging to exactly one cluster. However, there are often less crisp definite boundaries between clusters in real data sets so that the fuzzy partitioning which allows data points belonging to more than one cluster is usually more suitable for real applications. Zadeh [30] discussed the need for fuzzy logic in dealing with the vagueness of data. Fuzzy control charts based on fuzzy set theory are well-documented in literature [31–35]; however, to our best knowledge, few researches utilize fuzzy clustering in CP detection even though fuzzy clustering has been widely studied and applied in many substantive areas (cf. [36,37]). Alaeddini et al. [38], Zarandi and Alaeddini [39], and Kazemi et al. [40] employed a hybrid fuzzy-statistical clustering (FSC) method to identify a CP in phase II processes. They used true values of in-control process parameters when executing FSC, but they are rarely known exactly in real applications. Furthermore, FSC can only detect one CP of a single quality characteristic, thus, FSC cannot detect simultaneous changes in mean and variance neither multiple CPs.

As the mixture likelihood approach is powerful in CP detection (cf. [29]) and fuzzy clustering with fuzzy c-partitions is better suited for real applications, we consider transferring a CP model framework into a mixture model and then create a fuzzy classification maximum likelihood CP (FCML-CP) algorithm by using Yang's [41] FCML procedures to detect the time of shifts in mean and variance and simultaneously estimate shifts in different segments in which multiple changes are allowed. The proposed method is applicable to phase I and II processes without knowledge of in-control process parameters, and it can locate multiple CPs in a process at once.

The remainder of this paper is organized as follows. The fuzzy classification maximum likelihood procedure is reviewed briefly in Section 2. Then, a fuzzy classification maximum likelihood change-point (FCML-CP) algorithm for normal processes are developed in Section 3. In Section 4, extensive experiments with numerical and real data sets are conducted to