Computers & Industrial Engineering 125 (2018) 364–374

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Adaptive monitoring of multimodal data

Kai Wang[a], Jian Li[b,*], Fugee Tsung[a]

[a] Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[b] School of Management and State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, Shaanxi 710049, China

ARTICLE INFO

Keywords:
EWMA
Fused LASSO
Gaussian mixture model
Penalized likelihood
Statistical process control

ABSTRACT

Multimodal process data that include several subpopulations appear frequently in many complex applications due to process heterogeneity. Different from the most existing control charts that are only applicable to unimodal data, a new adaptive monitoring method is proposed in this paper for multimodal data from heterogeneous processes. Specifically, a Gaussian mixture model is first employed for data modeling. Considering the number of subpopulations that may change in Phase II, a penalized likelihood function is devised to infer the true number of subpopulations by shrinking any insignificant or redundant Gaussian components. Our proposed control chart, is thus not only sensitive to process changes in subpopulation parameters, but also adaptive to changes in the number of subpopulations. A diagnostic procedure is also followed to classify the changes in multimodal data. The superiority of our chart is fully demonstrated through numerical Monte Carlo simulations and a real industrial example in the production process of a 3D printing nylon powder material.

1. Introduction

Statistical process control (SPC) is an important methodology for achieving process stability. Despite a rich body of SPC charts developed in the literature, most of these charts, e.g., $\bar{x}$ chart, are implemented with an assumption that the normal process data come from a single operating mode and thus follow a unimodal distribution which has only one peak in the probability density function (PDF) (see Fig. 1). Some real complex processes, however, can be a combination of multiple operating modes, i.e., one operating mode may switch to another mode during a process. As a result, the process data may follow a multimodal distribution which has more than one peaks (local maxima) in the PDF (see Fig. 1). These process data that are generated by a distribution with multiple peaks are called multimodal data in this paper.

The multimodal process data in fact consist of several mixed subpopulations, each of which corresponds to an operating mode in the process, and they are commonly seen in many applications. For example, Yu and Qin (2008) illustrated a stirred tank heater process where the monitored variables, say temperature, would follow different distributions under different operating conditions. A mechanical milling process with different levels of cutting depth along its cutting trajectory results in the multimodality in the distribution of cutting force signal (Grasso, Colosimo, Semeraro, & Pacella, 2015). In nanoparticle engineering in Park and Shrivastava (2014), a sample of particles grow through multiple modes and finally converge into different shapes like circle, triangle and rectangle. The last example, which motivates this work, is the production process of a 3D printing powder material (see Section 5). Due to multiple thermal stages in the process, the in-control (IC) powder sizes follow a bimodal distribution.

This paper focuses on Phase II monitoring of multimodal data. Given a random sample $X_1, \ldots, X_n$, our purpose is to determine whether the current process distribution, in terms of the PDF f, has shifted or not. This parallels the hypothesis testing:

$$H_0: f = f^{(0)}, \quad H_1: f \neq f^{(0)}, \tag{1}$$

where $f^{(0)}$ is the IC distribution. For multimodal data with more than one subpopulation, H_1 can represent two classes of process changes. One class considers shifts in parameters associated with existing subpopulations in f, the other involves changes in the number of subpopulations. In view of this, a control chart for multimodal data should have capability in detecting both of these out-of-control (OC) types.

The control charts explicitly designed for multimodal data are still limited in the literature. Yu and Qin (2008), Ge and Song (2009) and Xie and Shi (2012) have studied the monitoring of multimode processes in chemical engineering. However, they all assume that the number of subpopulations which is known or well estimated from historical data is fixed in online monitoring. Therefore, their charts are only capable for detecting process shifts related to the existing subpopulations. Another

* Corresponding author.
E-mail addresses: kwangai@connect.ust.hk (K. Wang), jianli@mail.xjtu.edu.cn (J. Li), season@ust.hk (F. Tsung).

https://doi.org/10.1016/j.cie.2018.09.001
Received 7 March 2018; Received in revised form 31 August 2018; Accepted 2 September 2018
Available online 04 September 2018

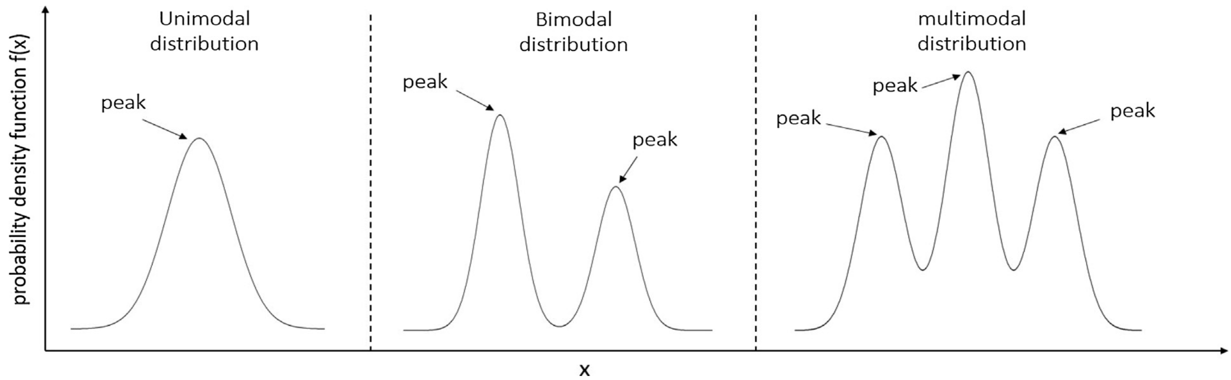

Fig. 1. Schematic plots of the PDFs of the unimodal, bimodal and multimodal distributions.

viable options are nonparametric or distribution-free charts which require no specific distribution forms. Although many related charts are developed based on the rank-related statistics (e.g., the Wilcoxon rank-sum statistic in Chong, Mukherjee, & Khoo (2017)) or goodness-of-fit tests (e.g., the Kolmogorov-Smirnov test in Ross & Adams (2012), the Cramér-von Mises test in Zhang, Li, & Li (2017) and the likelihood ratio-based test in Zou & Tsung (2010)), in effect, they all ignore multimodality information so that their charting performance may be hampered for multimodal data (see Section 4 for evidences).

To exploit multimodality, we adopt the Gaussian mixture model (GMM) in McLachlan and Peel (2000), Frühwirth-Schnatter (2006) and Kim, Lee, and Kim (2018) for its great flexibility and interpretability. It takes several Gaussian components, each of which represents one subpopulation in multimodal distributions. Based on the GMM, H_1 in the hypothesis testing (1) now indicates process changes in both the model parameters (the proportions and Gaussian component parameters) and the model order (the number of Gaussian components). Choi, Park, and Lee (2004), Yu and Qin (2008), Xie and Shi (2012) and Wang, Li, and Tsung (2018) also monitored the heterogeneous processes by GMMs, however, the number of Gaussian components in their models is assumed to be unchanged in Phase II.

This paper proposes a control chart that is adaptive to the number of subpopulations in multimodal data, so process changes in both the model parameters and the model order can be systematically monitored. By assuming the multimodal data come from a GMM, we design charting statistics by testing the hypotheses (1) via the likelihood ratio test (LRT). However, the fact that the model order may change in Phase II remains a major challenge. Specifically, the likelihood is non-decreasing over the model order of the GMM given a sample of data, as the GMM with more components has more model parameters. The likelihood alone thus cannot determine the model order. Many approaches have been proposed to overcome this difficulty, in which the model order is first determined using the Akaike information criterion (Akaike, 1998), Bayes information criterion (Schwarz, 1978), or distance measure (Chen & Kalbfleisch, 1996). The model parameters are then estimated accordingly. This sequential manner, however, has to search within a predefined model order range and is thus computationally intensive in online monitoring.

Actually, when fitting data by the GMM with a model order larger than the true value, two types of overfitting occur as in Chen and Khalili (2008). Type I overfitting has insignificant Gaussian components with near-zero proportions, whereas type II overfitting includes redundant Gaussian components of which the location parameters are very close. This discovery encourages penalties on the insignificant and redundant Gaussian components so that the model order and model parameters can be estimated in one strike from the likelihood function. Our chart is based on such penalized likelihood functions (see Section 2 for details), and is combined with the exponentially weighted moving average (EWMA) scheme.

This paper proposes a penalized LRT-based EWMA chart for monitoring multimodal data, which is a systematic tool capable of detecting process changes in both the model parameters and the model order. A diagnostic procedure is also attached for post-signal analysis. In the following, the adaptive modeling of multimodal data via a penalized likelihood and design of control chart are discussed in Sections 2 and 3. The numerical Monte Carlo simulations in Section 4 and the study of a real motivating example in Section 5 demonstrate the superiority of the proposed chart. Section 6 concludes this work. Some technical details are provided in the Appendix.

2. Adaptive modeling of multimodal data

As introduced in Section 1, our main challenge is modeling the multimodal data with uncertainty in the number of subpopulations in Phase II, and this is resolved by the penalized likelihood function in this section.

At first, we adopt the GMM to describe the multimodal data. Technically, the GMM represents the underlying mixed subpopulations through multiple Gaussian components, each of which has a proportion parameter, a mean and a variance parameter. Suppose X is a univariate random variable that follows a GMM with K components and model parameter $\mathbf{\Omega}_K$. Then its PDF is

$$f(x|\mathbf{\Omega}_K) = \sum_{k=1}^{K} p_k g(x|\theta_k) = \sum_{k=1}^{K} p_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\theta_k)^2}{2\sigma^2}\right),$$

where p_k is the proportion parameter of the kth component with $\sum_{k=1}^{K} p_k = 1$, $g(x|\theta_k)$ is the kth Gaussian component with location parameter θ_k, and the model parameter $\mathbf{\Omega}_K = \{\mathbf{p}_K, \boldsymbol{\theta}_K\} = \{p_1, \ldots, p_K, \theta_1, \ldots, \theta_K\}$.

Give a sample data $x_1, \ldots, x_n$, the above GMM degenerates to the Gaussian/normal distribution when $K = 1$ and $p_k = 1$, and tends to be the kernel density estimation when K continuously increases to be the sample size n with $p_k = 1/n$, $\theta_k = x_k$, the standard Gaussian distribution being the kernel and σ being the bandwidth parameter (McLachlan & Peel, 2000). Note that all Gaussian components here are assumed to share a same variance σ^2, since a finite GMM with a common variance and a large K already has sufficient flexibility to fit multimodal data in many applications (see Chen & Chen (2003), Chen & Kalbfleisch (2005) and Chen & Khalili (2008) for instance), and the common variance constraint can avoid unbounded likelihoods and spurious maximizers in parameter estimation (Hathaway, 1985). Therefore, the variance parameter is not the main concern of this paper and is assumed to be a known constant. The extension of our monitoring method to the case where the variance is also subject to changes in Phase II is explored in Section 4.3.

Since the model order K may change and is unknown in Phase II, we model the sample data $\mathbf{x} = (x_1, \ldots, x_n)^T$ using the GMM with a large value of K (upper bound). Then the log-likelihood function is