Int. J. Production Economics 141 (2013) 377–387

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

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

Minimal Euclidean distance chart based on support vector regression for monitoring mean shifts of auto-correlated processes

Shichang Du [a,]*, Jun Lv [b]

[a] Department of Industrial Engineering and Logistics Engineering, School of Mechanical Engineering, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, PR China

[b] School of Business, East China Normal University, Shanghai 200241, PR China

ARTICLE INFO

Article history:
Received 15 September 2011
Accepted 28 August 2012
Available online 15 September 2012

Keywords:
Minimal Euclidean distance
Support vector regression
Control chart
Statistical process control

ABSTRACT

Though traditional control charts have been widely used as effective tools in statistical process control (SPC), they are not applicable in many industrial applications where the process variables are highly auto-correlated. In this study, one new minimal Euclidean distance (MED) based monitoring approach is proposed for enhancing the monitoring mean shifts of auto-correlated processes. Support vector regression (SVR) is used to predict the values of a variable in time series. Through calculating minimal Euclidean distance (MED) values over time series, a novel MED chart is developed for monitoring mean shifts, and it can provide a comprehensive and quantitative assessment for the current process state. The performance of the proposed MED control chart is evaluated based on average run length (ARL). Simulation experiments are conducted and one industrial case is illustrated to validate the effectiveness of the developed MED control chart. The analysis results indicate that the developed MED control chart is more effective than other control charts for small process mean shifts in auto-correlated processes, and it can be used as a promising tool for SPC.

1. Introduction

The ability to monitor and reduce process variation for cost reduction and quality improvement in industrial processes plays a critical role in the success of one enterprise in today's globally competitive marketplace (Montgomery, 2001; Wu et al., 2007; Du et al., 2008; Lee et al., 2012). Control charts have been widely used as effective tools in statistical process control (SPC) for monitoring the process variation in industrial applications. In particular, control charts for monitoring independent observations have been extensively investigated and applied (Alt, 1984; Mason et al., 1995; Aparisi and Haro, 2003; Yang and Rahim, 2005; Torng et al., 2009; Magalhãesa et al., 2009; Wu et al., 2009; Du and Xi, 2010; Ou et al., 2011; Costa and Machado, 2011). With tremendous growth of advanced automatic data inspection and measurement techniques during the past few years, the process variables are being collected automatically at higher rates, therefore, for many industrial applications, a basic statistical assumption of independence is often violated, i.e. the data collected at regular time intervals from the processes is serially auto-correlated (Montgomery and Friedman, 1989; Cook and Chiu, 1998).

Several attempts have been made to extend traditional SPC techniques to deal with auto-correlated processes. One of the most interesting approaches to SPC for auto-correlated processes was proposed by Alwan and Roberts (1988). They introduced two charts, which they referred to as the common-cause control chart (CCC) and the special-cause control chart (SCC). CCC is a plot of forecasted values that are determined by fitting the correlated process with an autoregressive moving average model (ARMA), and SCC is a traditional Shewhart chart of the residuals. Their work attracted further investigation on time series modeling techniques application for monitoring correlated processes (Montgomery and Friedman, 1989; Montgomery and Mastrangelo, 1991; Wardell et al., 1992, 1994; Schmid, 1997; Adams and Tseng, 1998; Timmer et al.,1998; Jiang et al., 2000; Wright et al., 2001; Orlando et al., 2002; Kalgonda and Kulkarni, 2004). The time series based control charts approaches essentially involve fitting an adequate time series model to the correlated process data and applying a traditional control chart to the stream of residuals from the time series model. All these control chart approaches have been shown to improve the monitoring performance in the presence of auto-correlation. However, these time series modeling techniques require that a strict model has been identified for the time series of process observations before residuals can be obtained (Hwarng, 2004), and their performance is not very good for monitoring small shifts (Wardell et al., 1994), and they require one to have some skill in time series analysis (Box et al., 1994). Therefore, some other control charts based on residual have been developed for enhancing the performance of monitoring correlated processes (Testik, 2005; Pan and Jarrett, 2007).

* Corresponding author. Tel.: +86 137 644 79720; fax: +86 21 644 80001.
E-mail address: lovbin@sjtu.edu.cn (S. Du).

0925-5273/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2012.09.002

Recently, some researchers tried to find alternative methods that allow less restrictive assumptions, more flexibility and adaptability to real data situations. Examples of such techniques are machine learning methods such as neural network (NN) and support vector machine (SVM). These techniques allow learning the specific structure directly from the data and can be applied without forcing any assumptions. Some authors have proposed NN approaches as effective tools for monitoring auto-correlated processes (Cook and Chiu, 1998; Cook et al., 2001; Zobel et al., 2004; Hwarng, 2005; Pacella and Semeraro, 2007; Jamal et al., 2007; Du and Xi, 2011).

Support vector machine (SVM) has recently become a new generation learning system based on recent advances on statistical learning theory for solving a variety of learning, classification and prediction problems (Cortes and Vapnik, 1995; Gunn, 1998; Cristianini and Shawe-Taylor, 2000; Deng and Yeh, 2011). SVMs calculate a separating hyperplane that maximizes the margin between data classes to produce good generalization abilities. The main difference between NNs and SVMs is in their risk minimization (Gunn, 1998). In case of SVMs, structural risk minimization principle is used to minimize an upper bound based on an expected risk, whereas in NNs, traditional empirical risk minimization is used to minimize the error in the training of data. The difference in risk minimization leads to a better generalization performance for SVMs than NNs (Gunn, 1998). Support vector regression (SVR) is an important extension of SVM and is a regression method by introduction of an alternative loss function (Vapnik, 1998).

The applications of SVM to monitor the process variation are spare. Chinnam (2002) demonstrated that SVM can be extremely effective in minimizing both type I and type II errors for detecting shifts in the auto-correlated processes, and performed as well or better than traditional Shewhart control charts and other machine learning methods. Sun and Tsung (2003) and Kumar et al. (2006) developed one kernel-distance based K-chart using support vector for monitoring the independent observations. Issam and Mohamed (2008) presented SVR based cumulative sum (CUSUM) control chart for auto-correlated process. In their approach, SVR is firstly calculated and then CUSUM is applied to the stream of residuals from SVR. Therefore, two methods including SVR and CUSUM need to be calculated in their approach.

In this paper, one new minimal Euclidean distance (MED) based control chart is developed as a promising tool for monitoring auto-correlated processes. SVR is used to predict the values of a variable in time series. By using MED, the quantization error is provided for quantifying the deviation degree of current process with in-control process state space. Depending on how far away the current process is deviating from the in-control process state, a quantitative assessment index can be obtained by calculation of the MED of the new measurement data to the SVR trained by in-control process datasets.

The rest of this paper is organized as follows. The SVR theory is reviewed briefly in Section 2. The methodology using the proposed MED chart for monitoring auto-correlated processes is developed in Section 3. Experiments and performance analysis of the MED chart are conducted based on average run length (ARL) in Section 4. Further analysis of MED chart and one industrial application is illustrated to validate the effectiveness of the MED chart in Sections 5 and 6, respectively. A procedure for applying the MED chart into real processes is presented in Section 7. Finally, the conclusions and future work are given in Section 8.

2. Brief review of SVR

This section provides a brief review of SVR formulation. Consider a training data set $\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_N,y_N)\}$ $i=1,2,\ldots,N$, where N is the total number of training vectors, $\mathbf{x}_i \in R^d \subset R$ is the ith d-dimensional input vector, and $y_i \in \{1,-1\}$ is known target. The standard form of support vector regression (Vapnik, 1998) is:

$$\min_{\mathbf{w},b,\xi,\xi*} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{N}\xi_i+C\sum_{i=1}^{N}\xi_i^* \tag{1}$$

$$\text{subject to } \begin{cases} \mathbf{w}^T\phi(x_i)+b-y_i \le \varepsilon+\xi_i \\ y_i-\mathbf{w}^T\phi(x_i)-b \le \varepsilon+\xi_i^* \\ \xi_i,\xi_i^* \ge 0 \end{cases} \tag{2}$$

where $\mathbf{w}$ is the vector of hyperplane coefficients, defining a direction perpendicular to the hyperplane, the index i labels the N training cases, ξ_i and ξ_i^* are slack variables, measuring the degree of misclassification of the sample $\mathbf{x}_i$, C is the error penalty factor, penalizing the non-zero ξ_i, the bias b is a scalar, representing the bias of the hyperplane, the parameter ε is the ε-insensitive loss, and the map function ϕ is a non-linear transformation to map the input vectors into a high-dimensional feature space.

Eq. (2) corresponds to dealing with a called ε-insensitive loss function $|\xi|_\varepsilon$, which is one of most important loss functions (Vapnik, 1998). After transforming the problem to the dual form using Lagrangian transformation and applying optimality conditions, the following optimization problem is obtained:

$$\min_{\alpha,\alpha*} \frac{1}{2}(\alpha-\alpha^*)^T Q(\alpha-\alpha^*)+\varepsilon\sum_{i=1}^{N}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{N}y_i(\alpha_i-\alpha_i^*) \tag{3}$$

$$\text{subject to } \sum_{i=1}^{N}(\alpha_i-\alpha_i^*)=0, \quad 0\le\alpha_i,\ \alpha_i^*\le C \tag{4}$$

where $Q_{ij}=K(x_i,x_j)=\phi(x_i)^T\phi(x_j)$, and α_i and α_i^* are the Lagrange multipliers.

The estimated approximate function $\hat{f}$ is:

$$\hat{f}_{w,b}(x)=\sum_{i=1}^{N}(\hat{\alpha}_i-\hat{\alpha}_i^*)K(x_i,x)+b \tag{5}$$

where $K(x_i,x)$ is kernel function.

Any function that satisfies Mercer's theorem (Cristianini and Shawe-Taylor, 2000) can be used as a kernel function. Some popular SVM kernel functions include:

$$\text{linear function}: \quad K(\mathbf{x}_i,\mathbf{y}_j)=\mathbf{x}_i\times\mathbf{y}_j \tag{6}$$

Gaussian radial basis function (GRBF) :

$$K(\mathbf{x}_i,\ \mathbf{y}_j)=\exp(-\gamma||\mathbf{x}_i-\mathbf{y}_j||^2) \tag{7}$$

polynomial function with degree d :

$$K(\mathbf{x}_i,\mathbf{y}_j)=((\mathbf{x}_i\times\mathbf{y}_j)+\tau)^d \quad \tau\succ 0 \tag{8}$$

3. Methodology

Among various out-of-control conditions, this study is concerned with process mean shifts, which are defined as unanticipated sudden shifts in process mean vector. The primary possible causes for the mean shifts result from the introduction of new workers, machines or methods, a change in the measurement method or standard, etc.

3.1. MED chart based on SVR

In general, the in-control operation datasets are relatively easier to acquire, but it is hard to obtain lots of out-of-control datasets. Out-of-control state detection can be implemented