DETECTING THE PROCESS CHANGE FOR NONLINEAR PROFILE DATA

Jeh-Nan Pan, Department of Statistics, National Cheng Kung University, Tainan 70101, Taiwan, ROC, jnpan@mail.ncku.edu.tw
Chung-I Li, Department of Statistics, National Cheng Kung University, Tainan 70101, Taiwan, ROC, chunyilee1@gmail.com

ABSTACT

In this paper, the functional relationship of profile data is described via a non-parametric regression model and a nonparametric exponentially weighted moving average (EWMA) control chart is developed for detecting the process change for nonlinear profile data. We first fit the nonlinear profile data via a support vector regression (SVR) model and the nonparametric EWMA control chart with the five metrics can be constructed accordingly. Moreover, a simulation study is conducted to evaluate the detecting performance of the new chart under various process shifts. Finally, a realistic example is used to demonstrate the usefulness of the new chart and its monitoring schemes.

Keywords: Support vector regression (SVR), Nonparametric exponentially weighted moving average (EWMA) control chart, Nonlinear profile data, Metrics.

INTRODUCTION

Generally Speaking, the quality of a process or a product characterized by a functional relationship between the response variable (y) and one or more explanatory variables (x) is usually referred to as a profile. The process changes occurred in a functional relationship of the profile can be detected and classified by the profile monitoring and control. During the past decade, different methods have been proposed for profile monitoring in both Phases I and II. In the Phase I study, the parameters of the process are estimated based on a set of historical data and used to establish control limits for Phase II monitoring. In the Phase II, the data are sequentially collected over time to assess whether the parameters of the process have changed from the estimated values in the Phase I study. In this paper, we focus our research on the Phase II study for timely detecting the shifts/changes in the process parameters. Due to the complexity of parameters estimation involved in non-linear profile monitoring, Williams *et al.* [7] considered the use of nonparametric monitoring methods and the use of metrics to measure deviations from a baseline profile. After applying spline smoothing techniques to model the vertical density profile (VDP) data, they calculated the five metrics and employed an individual Shewhart control chart based on the moving range (I-MR chart) to establish control limits. However, according to Montgomery [4] and Pan *et al.* [5], exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts are proved to be more efficient than Shewhart control chart for detecting the small sustained shift in process mean. Thus, instead of using I-MR chart, we propose a revised non-parametric EWMA control chart in Phase II monitoring. Moreover, a more efficient support vector regression (SVR) method is used to detect and classify the process shifts in nonlinear profiles.

LITERATURE REVIEW

Non-linear profile monitoring were considered by other researchers. Williams *et al.* [7] proposed using three different T^2 statistics for Phase I analysis to monitor the coefficients resulting from a parametric nonlinear regression model that was used to fit profile data. They also considered the use of nonparametric regression method and the use of metrics to measure deviations from a reference profile. Hung *et al.* [2] proposed using support vector regression (SVR) to fit in-control profiles. Then, they employed the moving block bootstrap method to generate correlated samples for each in-control profile and obtain a simultaneous confidence region for the underlying functional relationship. The obtained confidence region was used to monitor the real AIDS data collected from hospitals in Taiwan. In practice, the functional relationship of the profile data rarely occurs in linear form and the real data usually do not follow normal distribution. Thus, in this paper, the functional relationship of profile data is described via a non-parametric regression model and a nonparametric EWMA control chart is developed for detecting the process change for nonlinear profile data in Phase II monitoring.

RESEARCH METHODOLOGY

Let y_{ij} be the measurement of the ith observations in the jth profile and x_{ij} be the corresponding explanatory variables such that $i = 1,2,\ldots,n_j$ for each $j = 1,2,\ldots$. When the process is in statistical control, the underlying model is assumed to be

$$y_{ij} = f_j\left(x_{ij}\right) + \varepsilon_{ij}, i = 1,2,\ldots,n_j, j = 1,2,\ldots \quad (1)$$

where $f(x)$ is a function with certain degree of smoothness, the random errors ε_{ij} are generally assumed to follow some distributions and the n_j's are taken to be equal and the explanatory variables are assumed to be fixed for different j's. According to Zou *et al.* [8], this is the common case for practical calibration applications in manufacturing industries.

In this paper, we use support vector regression (SVR) to fit profiles for describing the functional relationship shown in equation (1). The SVR is a supervised statistical learning algorithm for regression problem. In SVR, the explanatory variables are mapped onto a feature space, and a linear model in equation (2) is constructed in the feature space.

$$g\left(\mathbf{X},\mathbf{w}\right) = \mathbf{w}^T \phi\left(\mathbf{X}\right) + b \quad (2)$$

where $\mathbf{w}$ is normal vector, $\phi(.)$ is a nonlinear transformation function and $\mathbf{b}$ is bias term. The quality of estimation is measured by the loss function $L(y_i, g(\mathbf{X}_i, \mathbf{w}))$ and the SVR model is formulated as a minimization problem as listed in equation (3).

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (3)$$

subject to:

$$\begin{cases} y_i - g(\mathbf{X}_i, \mathbf{w}) \le \epsilon + \xi_i \\ g(\mathbf{X}_i, \mathbf{w}) - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (4)$$

where $\epsilon > 0$ is a certain threshold, constant $C > 0$ is a penalization that can be viewed as a way to control over-fitting and, ξ_i and ξ_i^* are slack variables.

The loss function is given by

$$L\big(y_i, g(x_i, \mathbf{w})\big) = \begin{cases} 0 & , \text{ if } |y_i - g(x_i, \mathbf{w})| < \epsilon \\ |y_i - g(x_i, \mathbf{w})| - \epsilon & , \text{ otherwise} \end{cases} \quad (5)$$

The eps-regression machine in R package e1071 (Meyer et al. [3]) is used to build the above SVR model based on the training dataset collected from the Phase I study. Through the SVR model, the baseline profiles are calculated as $\tilde{y}_i = \sum_{j=1}^{m} \hat{y}_{ij}/m$, $i = 1,\ldots,n$, where $\hat{y}_{ij}$ is the predicted values for the corresponding explanatory variables x_{ij}. In the Phase II study, let y_{ij} be the measurement of the ith observations in the jth profile and x_{ij} be the corresponding explanatory variables. The predicted values $\hat{y}_{ij}$ for the measurement y_{ij} are calculated based on the SVR model. Then, the five metrics suggested by Williams et al. [7] can be calculated as: $M_{j1} = \text{sign}\left(\max_{i=1,\dots,n}\{|\hat{y}_{ij} - \tilde{y}_i|\}\right)$, $M_{j2} = \sum_{i=1}^{n}|\hat{y}_{ij} - \tilde{y}_i|$, $M_{j3} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_{ij} - \tilde{y}_i|$, $M_{j4} = |M_{j1}|$ and $M_{j5} = \sum_{i=1}^{n}(\hat{y}_{ij} - \tilde{y}_i)^2$. In conducting the Phase II study, we adopt the concept of residual control chart and employ a nonparametric EWMA control chart proposed by Hackel and Ledolter [1] to monitor these metrics. In other words, we consider the metric for profile j as an individual observation and use a nonparametric EWMA control chart is to monitor the mean change of these five metrics with n observations over j time period, where M_1 denotes the maximum deviation, M_2 denotes the sum of absolute deviations, M_3 denotes the mean absolute deviation, M_4 denotes the absolute value of maximum deviation (disregarding the direction of dissimilarity), M_5 denotes the sum of squared differences between the predicted values for the measurement y_{ij} and the corresponding baseline profiles.

Let $\{M^*_{1,p}, M^*_{2,p},, \dots, M^*_{g-1,p}\}$ be a reference metric with size $g-1$, $p = 1, \dots, 5$, which are calculated from the Phase I in-control data. Let $M_{j,p}$, $j = 1,2, \dots$, be the pth metric for the corresponding jth profile in the Phase II study. The rank of $M_{j,p}$ with respect to the reference metric $\{M^*_{1,p}, M^*_{2,p},, \dots, M^*_{g-1,p}\}$ can be calculated as $R^*_{j,p} = 1 + \sum_{k=1}^{g-1} I(M_{j,p} > M^*_{k,p})$, where the indicator function $I(M_{j,p} > M^*_{k,p}) = 1$ if $M_{j,p} > M^*_{k,p}$, otherwise the indicator function = 0, and $g-1$ is the sample size of reference metric. Then, the standardized rank of $R^*_{j,p}$ can be calculated as

$$R_{j,p} = \frac{2}{g}\left(R^*_{j,p} - \frac{g+1}{2}\right) \quad (6)$$

Replacing the individual observation by the standardized rank $R_{j,p}$, our proposed EWMA statistics can be written as

$$\text{EWMA}_{j,p} = (1-\lambda)\text{EWMA}_{j-1,p} + \lambda R_{j,p}, j = 1, 2, \dots; \; 0 < \lambda \le 1 \quad (7)$$

where $\text{EWMA}_{0,p} = 0$ and λ is a smoothing parameter. Because the standardized ranks in equation (6) follow a discrete uniform distribution on the points $\{\frac{1}{g} - 1, \frac{3}{g} - 1, \dots, 1 - \frac{1}{g}\}$ with mean zero and variance $\frac{g^2-1}{3g^2}$, the control limits of our proposed nonparametric EWMA control chart is defined as

$$\pm L_R\sqrt{\frac{\lambda}{2-\lambda}\frac{g^2-1}{3g^2}} \quad (8)$$

where the width of the control limits L_R can be determined by Monte Carlo simulation for achieving a specified in-control average run length (ARL_0).

THE SIMULATION STUDY

In the simulation study, once the in-control average run length (ARL_0) is fixed at 370, the out-of-control average run length (ARL_1) is used to evaluate the detecting performance of our proposed nonparametric EWMA control chart after conducting 5000 simulation runs. As the exponential form is frequently used in relating y to x in nonlinear profile data, we assume that the in-control non-linear profile model follows an exponential form:

$$y_{ij} = 0.9\exp(\beta_0 + \beta_1 x_{ij}) + \varepsilon_{ij} \quad (9)$$

where $\beta_0 = 0.5$, $\beta_1 = 2$ and the explanatory variable are fixed as $x_{ij} = 0.01i$, $i = 1, \ldots, 300$ for each profile j. For error random terms ε_{ij}, we consider the following two scenarios: (i) a standard normal distribution $N(0, \sigma^2)$ with parameter $\sigma = 1$; (ii) an exponential distribution with rate parameter $\theta = 0.5$. Since a reference metric $\{M^*_{1,p}, M^*_{2,p}, , \ldots, M^*_{g-1,p}\}$ is needed to construct a Phase II control chart, a SVR model for obtaining the reference metric is built based on Phase I profile training data generated from the model shown in equation (3). In this paper, the sample size of the reference metric is set as $g = 21$. The simulation results for various control parameter L_R values under different λ are summarized in Table 1.

TABLE 1. Various control parameter L_R values under different λ with $g = 21$ by fixing ARL_0=370.

λ	0.05	0.1	0.2	0.3	0.4
L_R	1.37	1.77	2.10	2.18	2.17
λ	0.5	0.6	0.7	0.8	0.9
L_R	2.13	2.05	1.96	1.87	1.77

For the random error terms follow a standard normal distribution or an exponential distribution, the simulation results of ARL_1 values with $\lambda = 0.05$ and $\lambda = 0.1$ under different shifts in parameter β_0, β_1 and standard deviation σ can be obtained respectively. Note that the shifts in parameters β_0 and β_1 are defined as $\beta_0 + \delta$ and $\beta_1 + \delta$. And the shifts in parameters σ and θ are defined as $1 \times \delta$ and $0.5 + \delta$ for a standard normal and an exponential distribution respectively. The simulation results show that the detecting ability of the proposed nonparametric EWMA chart increases when the magnitudes of shift in parameter δ increases or the smoothing parameter λ decreases. In most cases, the ARL_1 of the nonparametric EWMA control chart with the metric M_1 and M_4 for detecting the shifts in various parameters are larger than those of M_2, M_3 and M_5, i.e. the detecting performances of the nonparametric EWMA control chart with metrics M_1 and M_4 are worse than those of M_2, M_3 and M_5 when the random error terms follow a standard normal distribution. Similar results can be obtained when the random error terms follow an exponential distribution.

NUMERICAL EXAMPLE

In order to demonstrate the practical application of our proposed control chart, the numerical example with vertical density profile (VDP) data given in Walker and Wright [6] is used for illustration purpose. In manufacturing the particle board, the density property of the finished board is the quality characteristic needs to be closely monitored. Each resulting profile consists of 314 density measurements and the distance between two consecutive measures is 0.002 inch. The original VDP data with 24 profiles are illustrated in Figure 1, where the depth of thickness of the particle boards ranges from 0 to 0.626 inch.

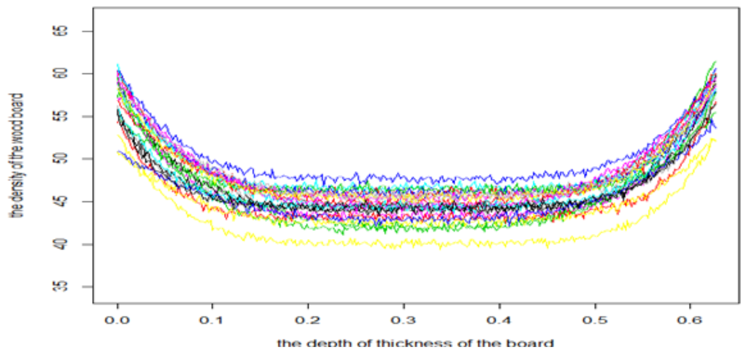

FIGURE 1. The original VDP data with 24 particle boards.

To construct a Phase II control chart, we first employ the SVR technique as mentioned in Research Methodology section to model the VDP data. The mean squared error (MSE) and mean absolute percent error (MAPE) for VDP data are summarized in Table 2.

TABLE 2. The MSE & MAPE for VDP data.

Profile	MSE	MAPE	Profile	MSE	MAPE
1	0.0467	0.48%	13	0.0649	0.48%
2	0.0635	0.53%	14	0.0601	0.53%
3	0.0654	0.48%	15	0.0469	0.48%
4	0.0589	0.50%	16	0.0410	0.50%
5	0.0610	0.48%	17	0.0505	0.48%
6	0.0577	0.49%	18	0.0630	0.49%
7	0.0529	0.53%	19	0.0656	0.53%
8	0.0525	0.49%	20	0.0501	0.49%
9	0.0498	0.47%	21	0.0547	0.47%
10	0.0471	0.48%	22	0.0533	0.48%
11	0.0617	0.47%	23	0.0513	0.47%
12	0.0509	0.49%	24	0.0533	0.49%

According to Table 2, the average MAPE can be approximated to 0.49%, which indicates its percent accuracy = 99.51%. Thus, we conclude that the fitted SVR model for VDP data is adequate. The 24 reference profiles obtained from the fitted model are shown in Figure 2.

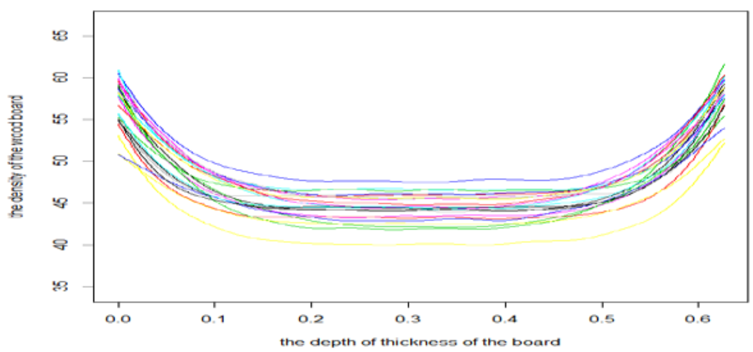

FIGURE 2. The 24 reference profiles after using SVR model to fit VDP data.

Since the predicted values can be calculated from the SVR model, the corresponding values of five metrics suggested by Williams *et al.* [7] are obtained accordingly. Then, the nonparametric EWMA control chart with five metrics is proposed to determine which board is out-of-control. Its upper and lower control limits are calculated based on equation (8) with $\lambda = 0.05$, $g = 21$ and $L_R = 1.37$. In Figure 3, the nonparametric EWMA control chart with five metric curves is plotted against the control limits according to their sample IDs. Figure 3 shows that all the VDP profiles using metric curves M1 and M4 are in-control. This result is consistent with the findings of Williams *et al.* [7]. It is worthy to note that Williams *et al.* [7] also pointed out the third profile is consistently higher than all other profiles. But, the nonparametric I-MR control chart proposed by Williams *et al.* [7] could not detect this an abnormal profile until the 6^{th} profile, whereas our proposed nonparametric EWMA control chart with the metric M2, M3 or M5 signals an out-of-control at the 4^{th} profile and these metric curves have the similar patterns. Apparently, the detecting ability of our proposed EWMA control chart outperforms the I-MR control chart proposed by Williams *et al.* [7].

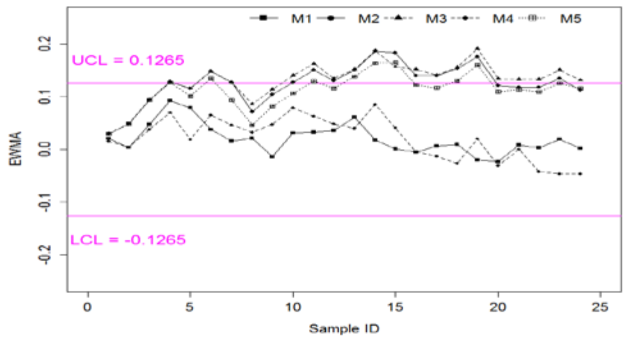

FIGURE 3. The nonparametric EWMA control chart with five metric curves: M1 denotes the maximum deviation; M2 denotes the maximum absolute deviation; M4 denotes the sum of squared differences; M5 denotes the mean absolute deviation.

CONCLUSIONS

In this paper, we first employ SVR model to smooth out the nonlinear profile data (i.e. to remove the noise of data). Then, the five metrics proposed by Williams *et al.* [7] can be obtained and the nonparametric EWMA control chart with five metrics is used to monitor the quality characteristic of the nonlinear profile data. Both the simulation results and numerical example show that our proposed nonparametric EWMA control chart with the metrics M2, M3 or M5 have better detecting performances than those of the I-MR control charts. Hence, the monitoring efficiency for nonlinear profile data in the Phase II study can significantly be enhanced by using our proposed nonparametric EWMA control chart with the metrics M2, M3 or M5. Hopefully, the results of this research can serve as a useful reference for quality practitioners when monitoring and controlling the process quality for the nonlinear profile data in the Phase II study.

REFERENCES

[1] Hackel, P., & Ledolter, J., A Control Chart Based on Ranks. *Journal of Quality Technology*, 1991, 23(2), 117-124.

[2] Hung, Y. C., Yang, S. F., Chuang, S. C., & Tseng, Y. K., Nonparametric Profile Monitoring in Multi-dimensional Data Spaces. *Journal of Process Control*, 2012, 22(2), 397-403.

[3] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2014). e1071., *Misc Functions of the Department of Statistics (e1071)*, TU Wien. R package version 1.6-3.

[4] Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*, 6^{th} Ed., Wiley, NY, 2009.

[5] Pan, J. N., Li, C. I., & Wu, J. J., A New Approach to Detecting the Process Changes for Multistage Systems. *Expert Systems with Applications*, 2016, 62(11), 293-301.

[6] Walker, E., & Wright S. P., Comparing Curves Using Additive Models. *Journal of Quality Technology*, 2002, 34(1), 118-129.

[7] Williams, J. D., Woodall, W. H., & Brich, J. B., Statistical Monitoring of Nonlinear Product and Process Quality Profiles. *Quality and Reliability Engineering International*, 2007, 23(7), 925-941.

[8] Zou, C., Tsung, F., & Wang, Z., Monitoring profiles based on nonparametric regression methods. *Technometrics*, 2008, 50(4), 512-526.