Research on Kernel-Distance-Based AEWMA-t Control Method

Yatang Yang\textsuperscript{1*}, Fumin Chen\textsuperscript{1}, Yiyu Tian\textsuperscript{1}

\textsuperscript{1} State Key Laboratory for Manufacturing Systems Engineering, Xi’an Jiaotong University, Xi’an; 710049, China
\* yangyatang@stu.xjtu.edu.cn

Abstract. Aimed at the complex data correlation among multivariate quality characteristics of products, inability to obey the assumptions of traditional control methods, low data volume in small and medium batch production, and inaccurate parameter estimation in process control, an AEWMA-t control method based on kernel distance data is proposed. Firstly, based on the support vector data description algorithm, the hypersphere is trained by normal samples. Then the kernel distance from the sample to the center of the hypersphere is calculated and the kernel distance is normally converted. Finally, the process is controlled with AEWMA-t control method. Case analysis shows that, compared with the traditional multivariate control methods, this method has good ability to detect anomalies in the mean deviation interval of each process, and is not restricted by the distribution and size of the data samples.

1. Introduction

As the production process become increasingly complex, processes need to monitor two or more key quality characteristics. However, the hypothesis distribution between the parameters of multiple characteristics cannot be determined, so the traditional multiple statistical process control theory is not applicable. Besides, the size of data in medium and small batch production process is small. This leads to inaccurate estimation on relevant parameters of control chart. Therefore, real-time monitoring on multivariate quality characteristics in the medium and small batch production mode has become a research hotspot.

In order to solve the limitation of traditional multivariate quality control methods\cite{1}, such as T2 and MEWMA, which require joint normal distribution of feature data, Sun and Tsun et al. introduced the idea of machine learning and proposed a multivariate quality control method based on SVM\cite{2}. For the inaccurate estimation of mean value and variance in small and medium batch production processes, Zhang et al. proposed EWMA-t control method\cite{3} and improved the ability to detect small deviation of process. To some extent, the SVM idea can solve the problem of complex correlation between data, but it cannot detect the small deviation of process in time, and EWMA-t control method is only applicable to unary process control.

Based on the problems in the researches above, we propose a kernel-distance-based AEWMA-t control method (K-AEMWA-t). The rest of this paper is arranged as follows: the second part introduces the research framework; the third part introduces the process of building control method modeling; the fourth part provides an analyzing case; the fifth part includes summary and discussion.
2. Research Framework

The figure 1 illustrates the research process of K-AEWMA-t control method and this method is suitable for multivariate quality control. Its main steps are below: combining support vector data description algorithm with Gaussian kernel function to train the data collected in the early normal production process; choosing the high-quality relevant training parameters through particle swarm optimization algorithm, in order to obtain the normal-sample hypersphere model; taking the kernel distance D which from the new production process data to the core of the hypersphere—as the new monitoring point, and normally transforming it by BOX-COX algorithm; in the end, introducing adaptive filtering algorithm into EWMA-t method to realize the steady state control on the production process.

![Figure 1. Research process of AEWMA-t process control method based on kernel distance](image)

3. AEWMA-t Control Method Modeling Based on Kernel Distance

3.1. Support Vector Data Description Algorithm

Support Vector Data Description (SVDD) is a single-value classification method. Its theoretical idea is to train a minimum compact super sphere \( F \) with a center of \( a \) and a radius of \( R \), and to include as many described objects as possible[4]. Assuming \( X \) to be n target class training samples with m dimensions, then the hypersphere model can be defined as:

\[
\min F(R,a,\xi) = R^2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} \|X_i - a\|^2 \leq R^2 + \xi_i, i = 1,2,...,n, \xi_i \geq 0, i = 1,2,...,n
\]

where \( \xi_i \) is the relaxation variable; \( C \) is the penalty coefficient, the penalty degree of the wrong sample plays a system role.

Lagrange function is introduced to transform into an unconstrained optimal problem.

\[
L(R,a,\gamma,\xi) = R^2 + \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left( R^2 + \xi_i - \|X_i - a\|^2 \right) - \sum_{i=1}^{n} \gamma_i \xi_i
\]

where Lagrange multipliers \( \alpha_i \geq 0, \gamma_i \geq 0 \). Take the partial derivatives \( R, a \) and \( \xi_i \) of equation (2) and set them equal to 0, and turn this into the dual problem of equation (2).

\[
\max L = \sum_{i=1}^{n} \alpha_i \left( X_i \cdot X_i \right) - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \left( X_i \cdot X_j \right)
\]

In order to simplify the operation complexity and make the model boundary more compact, the gaussian kernel function \( k(x_i,x_j) \) is introduced to replace the inner product \( (X_i \cdot X_j) \) for calculation[5].

The expression of gaussian kernel function is:

\[
k(x_i,x_j) = \exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)
\]

Substitute gaussian kernel function into formula (3), then

\[
\max L = \sum_{i=1}^{n} \alpha_i Q_{xx} - \alpha^T Q \alpha ; \text{s.t.} \sum_{i=1}^{n} \alpha_i = 1,0 \leq \alpha_i \leq C, i = 1,2,...,n
\]
where $Q$ is the kernel matrix and the formal is transferred into the basic form of quadratic programming for solving, and the optimal solution set, $\mathbf{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, can be obtained, where the $x_i$ corresponding to $\alpha_i$, satisfying $0 \leq \alpha_i \leq C$, are the support vectors.

According to the principle of SVDD algorithm, the hypersphere distance is introduced and the kernel distance of the sample under test to the hypersphere is taken as the new monitoring point of production process control, then

$$D_k(z) = \sqrt{k(z, z) - 2 \sum_{i=1}^{n} \alpha_i k(x_i, z) + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j)}$$  \hspace{1cm} (6)

In formula (6), kernel function parameters $\sigma$ and penalty coefficient $C$ play a key role in the calculation of kernel distance $D_k$, so particle swarm optimization algorithm [6] is proposed to select the best hypersphere parameters.

3.2. BOX-COX Data Conversion Method

Multivariate quality characteristics in sample data, after dealing with the algorithm of support vector data description, are transferred into unary representation of kernel distance data. The production process can be monitored by a meta control, but because of its not obeying normal distribution, therefore, kernel distance data should be transferred by BOX-COX algorithm [7].

The transformation form of BOX-COX transformation analysis is as follows:

$$y' = \begin{cases} \left( y^k - 1 \right) / \lambda, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases}$$  \hspace{1cm} (7)

where $y'$ is the transformed sample data and $\lambda$ is the transformation parameter. For a given target $y = [y_1, y_2, \ldots, y_n]$ to be transformed, the optimal value can be obtained by maximizing the natural logarithm of the maximum likelihood probability function, then the targeted function is:

$$\max f(y, \lambda) = -\frac{n}{2} \ln \left[ \sum_{i=1}^{n} \left( y_i(\lambda) - \bar{y}(\lambda) \right)^2 / n \right] + (\lambda - 1) \sum_{i=1}^{n} \ln(y_i)$$  \hspace{1cm} (8)

where $\bar{y}(\lambda) = \frac{1}{n} \sum_{i=1}^{n} y_i(\lambda)$ is the arithmetic mean value of the transformed data.

In order to ensure that the converted data will not lose its original migration information when an anomaly occurs in the production process, the data collected under normal production state is used to obtain the conversion coefficient $\lambda$, and the subsequent detection data is converted based on $\lambda$.

3.3. AEWMA-t Control Chart Design Based on Kernel Distance Data

The control chart design is divided into control plot point design and control limit design. In the previous stage of data processing, the multivariate sample data has been transformed into unitary representational data subject to normal distribution. Suppose that each batch of samples after conversion is $X = \{X_1, X_2, \ldots, X_i\}$, and each batch of samples is $X_i = \{x_{i1}, x_{i2}, \ldots, x_{ij}\}$, where $i$ represents the batch of samples $i = 1, \ldots, m$; $J$ is the size of each batch of samples $j = 1, \ldots, n$.

(1) Control Variable Definition.

For the EWMA-t control method, the control variables $Z_i$ are defined as follows:

$$Z_i = (1-\lambda)Z_{i-1} + \lambda T_i$$  \hspace{1cm} (9)

where $Z_0 = 0$; $\lambda$ is the control limit smoothing coefficient $\lambda \in (0,1]$; $T_i$ is the value of the statistical variable defined according to the t control chart, then

$$T_i = \left( \overline{X}_i - u_0 \right) \sqrt{n / S_i}$$  \hspace{1cm} (10)

$\overline{X}_i$, $S_i$ are the mean value and variance of the ith batch sample respectively, then
The EWMA-t control method can effectively detect the small deviation in the process, but its inertia will lead to the delaying detection to the large deviation. Therefore, an adaptive filtering method [8] is introduced to construct an AEWMA-t control diagram suitable for each abnormal fluctuation range.

The control variable of equation (9) is transformed as follows, then

$$Z_t = (1-\lambda)Z_{t-1} + \lambda(T_t - Z_{t-1}) = Z_{t-1} + \phi(e_t)$$

(12)

where $e_t$ is the error between the current variable and the historical variable, and $e_t = T_t - Z_{t-1}$; $\phi(e_t)$ is a function of the error term $e_t$. It can be constructed as follows:

$$\phi(e_t) = \begin{cases} e_t + (1-\lambda)k, & e_t < -k \\ e_t - (1-\lambda)k, & e_t > k \\ \lambda e_t, & |e_t| \leq k \end{cases}$$

(13)

where $k$ is the threshold of deviation, $k > 0$. Set $\Delta(e_t) = \phi(e_t)/e_t$, substitute into equation (12)

$$Z_t = (1-\Delta(e_t))Z_{t-1} + \Delta(e_t)T_t$$

(14)

$\Delta(e_t)$ is equivalent to the smoothing coefficient $\lambda$. When there is a large deviation in the process, the error function $\phi(e_t)$ makes the equivalent smoothing coefficient $\Delta(e_t)$ become larger, and the ability to detect large deviation is enhanced, so as to realize the adaptive control smoothing coefficient in different migration ranges.

(2) Control Limit Design.

Due to the complexity of control variables, the control limit of K-AEWMA-t control method is preliminarily defined as:

$$UCL = h; LCL = -h$$

(15)

where $h$ is defined as the control limit obtained when the running chain length in the controlled state is a certain value, which can be calculated by Markov chain method [9].

4. Empirical case

Taking the machining process of blade profile of a certain type of turbine blade as an example, the production process was controlled based on the multiple quality characteristics of blade profile, and the performance was compared with the K-SVDD control chart proposed in literature [10]. At the same time, the influence of different sample sizes on the detection performance of K-AEWMA-t control chart was analyzed.

4.1. Establishment of Control Chart

From the 100 profile feature data collected in the early stage under normal processing state, multivariate quality data processing was carried out according to steps 3.1-3.2. Setting chain length of controlled operation $ARL_0=200$ in normal production state, using the data in normal samples and combing the constrained particle swarm optimization algorithm to select the parameters of hypersphere model. The adaptive function of particle swarm optimization algorithm is defined as:

$$f = g(\sigma, C) = C_{in}/(C_{in} + C_{out}); \quad f_{max} = 0.995$$

(16)

where the $C_{in}$ and $C_{out}$ is the number of normal test samples within and outside the radius of the hypersphere in the SVDD model obtained by substituting parameter training $\sigma, C$. 

\[
\begin{align*}
\bar{X}_i &= \frac{1}{n} \sum_{j=1}^{n} x_{ij} \\
S_i &= \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{X}_i)^2}
\end{align*}
\]

(11)
The process and results of particle swarm optimization for hypersphere model parameters are shown in figure 2. When the kernel function parameter is 0.01 and the penalty coefficient is 0.491, the fitness value is 0.995, and the radius $R$ obtained by SVDD training is 0.08751.

![PSO iterative process](image1)

(a) PSO iterative process

![3D cloud map with parameter selection](image2)

(b) 3D cloud map with parameter selection

Figure 2 Parameter selection of hypersphere based on PSO

4.2. Comparative Analysis

In order to compare the control performance of K-AEWMA-t and K-SVDD control charts, it is necessary to analyze the abnormal detection capability of different control charts when the mean value of the process is shifted at different times under the condition that the controlled running chain length (ARL0) is certain. According to 4.1, when ARL0=200, the control limit of the K-SVDD control chart can be defined as the radius of the hypersphere, then $H_{SVDD} = 0.0875$, therefore, under the condition of ARL0=200, the performance of the k-AEWMA-t and k-SVDD control chart can be compared and analyzed. The control limit of K-AEWMA-t can be obtained by MARKOV chain calculation method, that is, when $\lambda = 0.094, k = 8.82$ and UCL=0.7908, ARL0=200.15.

On the basis of normal samples, 300 sets of abnormal data were generated from the mean value of samples according to the deviation $\delta = [0.25, 0.5, 1, 1.5, 2, 2.5]$, and the control method mentioned in simulation and K-SVDD control method were simulated to analyze the out-of-control running chain length (ARL1) under different process offsets. The comparative analysis results are shown in figure 3, and the specific data are shown in table 1.

![ARL and $\delta$](image3)

Figure 3. Two different control methods for ARL1

| $\delta$ | 0     | 0.25  | 0.5   | 1     | 1.5   | 2     | 2.5   |
|---------|-------|-------|-------|-------|-------|-------|-------|
| K-SVDD  | 200.00| 28.731| 18.867| 5.514 | 2.195 | 1.226 | 1.007 |
| K-AEWMA-t| 200.15| 25.697| 14.642| 4.826 | 2.264 | 1.302 | 1.122 |

Table 1. ARL1 for K-SVDD chart and improved AEWMA-t chart

It can be seen from the comparison results that the abnormal detection output of the K-AEWMA-t control chart presented in this paper is significantly better than that of the K-SVDD control chart within the small and medium migration range, and the control performance of the two is very close within the large migration range. In order to verify the influence of different sample sizes on the
The performance of K-AEWMA-t control chart, ARL0=370.4 was taken as the standard, and the Markov chain method was used to calculate the ARL1 of control charts under different $\delta$ lower, when $n=[3,5,7]$. The analysis results and related control parameters are shown in figure 4 and table 2.

![Figure 4. Influence of sample size on the performance of K-AEWMA-t control chart](image)

Table 2. ARL1 for K-SVDD chart and K-AEWMA-t chart

| n  | UCL | $\lambda$ | $k$          | 0       | 0.25    | 0.5      | 1       | 1.5    | 2       | 2.5   | 3      |
|----|-----|----------|--------------|---------|---------|----------|---------|--------|---------|-------|--------|
| 3  | 0.693 | 0.026   | 37.39       | 370.417 | 67.765  | 26.248   | 11.691  | 7.718  | 5.880   | 4.818 | 4.125  |
| 5  | 0.910 | 0.094   | 8.82        | 370.444 | 36.339  | 11.606   | 4.843   | 3.193  | 2.454   | 2.025 | 1.747  |
| 7  | 0.691 | 0.098   | 4.53        | 370.412 | 35.269  | 10.196   | 3.992   | 2.301  | 1.519   | 1.171 | 1.044  |

As can be seen from table 2, when the sample size $n$ is 3, the control figure ARL1 is significantly larger than the situation when $n = 5$ or $7$. At the same time, when $n = 5$ or $7$, the difference on ARL1 is relatively small. Therefore, considering that the production process data volume is small in the medium and small batch production mode, and in order to obtain a smaller ARL1, the sample size $n=5$ is selected as the sample data volume detected in the production process.

References

[1] Faraz A, Cédric H, Saniga E M. (2012) Optimal T-2 Control Chart with a Double Sampling Scheme-An Alternative to the MEWMA Chart. Quality and Reliability Engineering, 28(7):751-760.

[2] Sun R, Tsung F. (2003) A kernel-distance-based multivariate control chart using support vector methods. International Journal of Production Research, 41(13):2975-2989.

[3] Zhang L, Chen G, Castagliola P. (2009) On t and EWMAt Charts for Monitoring Changes in the Process Mean. Quality and Reliability Engineering, 25(8):933-945.

[4] Li L J, Lei W P, Han J, et al. (2012) Intelligent Fault Diagnosis Method Based on Vector-Bispectrum and SVDD. Advanced Materials Research, 490-495:1029-1033.

[5] Prudhvi G, Heesung K. (2011) Support-Vector-Based Hyperspectral Anomaly Detection Using Optimized Kernel Parameters. IEEE Geoscience & Remote Sensing Letters, 8(6):1060-1064.

[6] Yanlong Z, Nan H, Zhenhua Y. Improved particle swarm algorithm based on secondary search. (2017) Journal of Computer Applications, 37 (9): 2541-2546.

[7] Bicego M, Baldo S. Properties of the Box-Cox Transformation for Pattern Classification. (2016) Neurocomputing, 218:390-400.

[8] Saleh N A. The Performance of the Adaptive Exponentially Weighted Moving Average Control Chart with Estimated Parameters. (2013) Quality and Reliability Engineering, 595-606.

[9] Zhiyuan C, Jinsheng S. Statistical economic design of adaptive EWMA control char. (2016) Control and Decision, 9: 1715-1719.

[10] Gani W, Limam M. Performance Evaluation of One-Class Classification-based Control Charts through an Industrial Application. (2013) Quality and Reliability Engineering, 29(6):841-854.