

Research Article

A One-Class Classification-Based Control Chart Using the K -Means Data Description Algorithm

Walid Gani¹ and Mohamed Limam²

¹ LARODEC, ISG, University of Tunis, 41 Avenue de la Liberté, 2000 Bardo, Tunisia

² Dhofar University, P.O. Box 2509, 211 Salalah, Oman

Correspondence should be addressed to Walid Gani; walid.gani@cct.gov.tn

Received 27 December 2013; Revised 23 April 2014; Accepted 7 May 2014; Published 9 June 2014

Academic Editor: Yi-Hung Chen

Copyright © 2014 W. Gani and M. Limam. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper aims to enlarge the family of one-class classification-based control charts, referred to as OC-charts, and extend their applications. We propose a new OC-chart using the K -means data description (KMDD) algorithm, referred to as KM-chart. The proposed KM-chart gives the minimum closed spherical boundary around the in-control process data. It measures the distance between the center of KMDD-based sphere and the new incoming sample to be monitored. Any sample having a distance greater than the radius of KMDD-based sphere is considered as an out-of-control sample. Phase I and II analysis of KM-chart was evaluated through a real industrial application. In a comparative study based on the average run length (ARL) criterion, KM-chart was compared with the kernel-distance based control chart, referred to as K-chart, and the k -nearest neighbor data description-based control chart, referred to as KNN-chart. Results revealed that, in terms of ARL, KM-chart performed better than KNN-chart in detecting small shifts in mean vector. Furthermore, the paper provides the MATLAB code for KM-chart, developed by the authors.

1. Introduction

In recent years, several attempts have been proposed to integrate data mining with statistical process control (SPC) [1–6]. The objective was to overcome the limitations of traditional parametric control charts especially the normality assumption, which may not be applicable in the case of modern manufacturing systems. The most commonly applied data mining technique in SPC is one-class classification. Actually, one-class classification methods have been widely used for process monitoring [7–12]. The principle of one-class consists in constructing a sphere which contains the maximum of data with minimum volume. This sphere distinguishes in-control process data, also known as target data, from out-of-control process data. The shape and volume of the one-class depend on the used one-class classifier, also known as data description algorithm. The one-class approach was applied to develop a new family of control charts called one-class classification-based control charts, referred to as OC-charts.

Several types of one-class classifiers exist in the literature. For instance, only two one-class classifiers were

used to develop control charts, named the support vector data description (SVDD) algorithm [13] and the k -nearest neighbor data description (KNNDD) algorithm [14]. Sun and Tsung [15] used SVDD to develop the kernel distance-based multivariate control chart, also known as the K-chart, which is considered as the first OC-chart that uses support vector principles. When monitoring more than two variables, the K-chart uses kernel methods that provide the advantage of dealing with high-dimension data. During the last decade, the K-chart has received significant attention and has witnessed several improvements through many works [16–19].

The SVDD algorithm was also employed to develop other control charts such as the SVDD based multivariate cumulative sum control chart [20], monitor batch process [21], monitor nonlinear processes [22], and perform industrial calibration [23]. Despite its successful application, the high computational cost of SVDD remains its main drawback. Actually, the SVDD algorithm loses its efficiency when the size of the training data becomes large. To overcome this shortcoming, Sukchotrat et al. [24] proposed the use of KNNDD algorithm to develop the KNN-chart. KNNDD

is a simple and fast algorithm that performs better with high dimensional data and does not consume much time during the training phase. Gani and Limam [25] compared the performance of the K-chart and the KNN-chart and demonstrated that the K-chart is sensitive to small shifts in mean vector, while the KNN-chart is sensitive to moderate shifts in mean vector.

This paper investigates the use of another one-class classifier which is the K -means data description (KMDD) algorithm to construct a new OC-chart, referred to as the KM-chart. The objective of this work is twofold. First, we aim to enlarge the family of OC-charts and extend their applications by showing the methodology of their construction and providing the necessary software codes. Second, we attempt to propose an OC-chart that can compete with K-chart and KNN-chart in terms of the average run length (ARL) criterion.

The rest of this paper is organized as follows. A review of OC-charts is presented in Section 2. The proposed KM-chart is introduced in Section 3. Construction methodology of the KM-chart using a real data example is shown in Section 4, while performance analysis of the proposed control chart is discussed in Section 5. Section 6 summarizes this paper.

2. Background on OC-Charts

In the literature, there are two common OC-charts which are the K-chart and the KNN-chart. In the following, we give a review of these control charts.

2.1. The K-Chart. The K-chart relies on SVDD algorithm, which an unsupervised one-class classifier, to fit a sphere around the target data. This sphere is determined by solving the following quadratic programming

$$\text{Minimize } F(R, a) = R^2, \quad (1)$$

subject to

$$\|x_i - a\|^2 \leq R^2, \quad i = 1, 2, \dots, n, \quad (2)$$

where F , a , and R , are, respectively, the cost function to minimize, the center and the radius of the sphere. Equation (2) shows that the vector of quality characteristics, denoted by x_i , having a distance smaller than the radius are considered as target. To allow the possibility of having outliers in the training set, the distance from x_i to the center a should not be strictly smaller than R^2 , and larger distances should be penalized. Therefore, we introduce slack variables $\xi_i \geq 0$ and the minimization problem becomes

$$\text{Minimize } F(R, a) = R^2 + C \sum_{i=1}^n \xi_i, \quad (3)$$

subject to

$$\|x_i - a\|^2 \leq R^2 + \xi_i, \quad (4)$$

where $C > 0$ is a parameter introduced for the trade-off between the volume of the sphere and the errors.

Equation (4) can be incorporated into (3) by using Lagrange multipliers

$$\begin{aligned} L(R, a, \alpha_i, \eta_i, \xi_i) = & R^2 + C \sum_{i=1}^n \xi_i \\ & - \sum_{i=1}^n x_i \left[R^2 + \xi_i - (\|x_i\|^2 - 2a \cdot x_i + \|a\|^2) \right] \\ & - \sum_{i=1}^n \eta_i \xi_i, \end{aligned} \quad (5)$$

with the Lagrange multipliers $\alpha_i \geq 0$ and $\eta_i \geq 0$, L should be minimized with respect to R , a , and ξ_i and maximized with respect to α_i and η_i . Setting partial derivatives of L , we obtain

$$\sum_{j=1}^m \alpha_j = 1, \quad (6)$$

$$a = \sum_{i=1}^n \alpha_i x_i, \quad (7)$$

$$C - \alpha_i - \eta_i = 0. \quad (8)$$

From (8), $\alpha_i = C - \eta_i$, $\alpha_i \geq 0$, and $\eta_i \geq 0$, then Lagrange multipliers η_i can be removed and we have

$$0 \leq \alpha_i \leq C. \quad (9)$$

By substituting (6) and (8) into (5), we have

$$\text{Maximize } L = \sum_{i=1}^n \alpha_i (x_i \cdot x_i) - \sum_{i,j=1}^n \alpha_i \alpha_j (x_i \cdot x_j), \quad (10)$$

subject to

$$0 \leq \alpha_i \leq C. \quad (11)$$

A test sample, denoted by z , is accepted when its distance is smaller or equal to the radius. This is equivalent to

$$\begin{aligned} (z - a)' (z - a) = & (z' \cdot z) - 2 \sum_{i=1}^n \alpha_i (z \cdot x_i) \\ & + \sum_{i,j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2. \end{aligned} \quad (12)$$

Generally, data is not spherically distributed. To make the method more flexible, the vectors of x_i are transformed to a higher dimensional feature space. The inner products in (10) and (12) are substituted by a kernel function $K(x_i, x_j)$. In a higher dimension, the sphere becomes a complex form called "hypersphere." The problem of finding the optimal hypersphere is given by

$$\text{Maximize } L = \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j), \quad (13)$$

subject to (11).

A test sample z is accepted when

$$K(z, z) - 2 \sum_{i=1}^n \alpha_i K(z, x_i) + \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j) \leq R^2. \quad (14)$$

The construction of the K-chart consists in determining which samples are support vectors (SVs) by solving the following quadratic programming

$$\text{Maximize } \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j), \quad (15)$$

subject to

$$0 \leq \alpha_i \leq C. \quad (16)$$

Once the SVs are obtained, the kernel distance (KD) of each sample is computed. For a test sample z , the KD is computed as follows

$$\text{KD}_z = \sqrt{K(z, z) - 2 \sum_{i \in S} \alpha_i K(z, x_i) + \sum_{i,j \in S} \alpha_i \alpha_j K(x_i, x_j)}, \quad (17)$$

where S is the set of SVs.

The KD of SVs, denoted by KD_{SVs} , represents the upper control limit (UCL) for the K-chart used to monitor a new sample z . This can be illustrated by the following hypothesis test

$$H_0 : \text{KD}_z \leq \text{KD}_{\text{SVs}} \text{ versus } H_1 : \text{KD}_z > \text{KD}_{\text{SVs}}. \quad (18)$$

Under H_0 the process is considered as in-control and under H_1 the process is considered as out-of-control, when sample z was taken.

2.2. The KNN-Chart. The KNN-chart uses an unsupervised one-class classifier called KNNDD to construct a one-class by estimating the local density of data. To understand the mechanism of the KNN-chart a brief description of the KNNDD algorithm, based on the work of Sukchotrat et al. [24], is presented below.

Let $NN_i(x)$ be the i th nearest neighbor training observation of data point x that needs to be monitored. Let V be the volume of the hypersphere containing i nearest neighbor training observations and N the size of the training set. The local density of x , denoted by $d(x)$, can be determined as

$$d(x) = \frac{i/N}{V(\|x - NN_i(x)\|)}. \quad (19)$$

Similarly, the local density of $NN_i(x)$, denoted by $d(NN_i(x))$, can be determined by

$$d(NN_i(x)) = \frac{i/N}{V(\|NN_i(x) - NN_i(NN_i(x))\|)}, \quad (20)$$

where $NN_i(NN_i(x))$ is the i th nearest neighbor of $NN_i(x)$ in the same training set.

The KNNDD method classifies x as the target class when the ratio of its local density of x is greater than or equal to one, which can be explained as follows

$$\frac{d(x)}{d(NN_i(x))} = \frac{\|NN_i(x) - NN_i(NN_i(x))\|}{\|x - NN_i(x)\|} \geq 1. \quad (21)$$

To make the algorithm more robust, the average of k -distances is considered (for $i = 1, \dots, k$). Thus, (21) becomes

$$\frac{\sum_{i=1}^k \|NN_i(x) - NN_i(NN_i(x))\|}{\sum_{i=1}^k \|x - NN_i(x)\|} \geq 1. \quad (22)$$

To construct the KNN-chart, the K^2 statistic representing the average distance between x and the k -nearest observations is computed as follows

$$K^2 = \frac{\sum_{i=1}^k \|x - NN_i(x)\|}{k}. \quad (23)$$

The K^2 values are used as monitoring statistics.

3. The Proposed KM-Chart

The proposed KM-chart gives the minimum closed spherical boundary around the in-control process data using the KMDD algorithm. The latter is an unsupervised one-class classifier, based on the K -means algorithm which is a very popular clustering method. It measures the distance between the KMDD-based sphere and the new incoming sample to be monitored. The sphere is described by K clusters placed such that the average distance to a cluster center is minimized.

The phase I of the KM-chart consists in determining the optimal KMDD-based one-class, by estimating the optimal number of clusters. In this step, the K -means clustering algorithm aims to find K clusters, denoted by C_1, \dots, C_K , that minimize the within clusters sum of squares as follows

$$D^2 = \min_{C_1, \dots, C_K} \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2, \quad i = 1, \dots, n, \quad k = 1, \dots, K, \quad (24)$$

where C_1, \dots, C_K are the disjoint sets of cluster indices, n is the number of observations in the training phase, μ_k is the sample mean of the observations in the k th cluster, and D^2 is the Euclidean distance of the quality characteristic x_i . It should be noted that the D^2 values are used as charting statistics for the KM-chart. The optimization problem in (24) can be solved by iterating the following two steps.

- (i) Given cluster centers μ_1, \dots, μ_K , assign each point to the cluster with the closest center.
- (ii) Given a cluster assignment, update the cluster centers to be the sample mean of the observations in each cluster.

In phase II, the distance of a new incoming sample, denoted by z_j , is computed as follows

$$D^2(z_j) = \min_{1, \dots, k} \|z_j - \mu_k\|^2, \quad j = 1, \dots, m, \quad k = 1, \dots, K, \quad (25)$$

where m is the number of observations in the testing phase.

```

%Let Q be the (n × m) matrix of quality variables where n is the number
%of observations and m the number of quality variables in the training phase.
%Define the target class.
T=target_class(+Q);
%Use the KMDD algorithm to fit a sphere around the defined target class above,
%where c1 is a fraction error on the target class and c2 is a parameter
%defining the number of clusters.
w = kmeans_dd(T,c1,c2);
%Show the results of KMDD classifier.
W=+w;
%Phase I of the KM-chart:
%Compute the Euclidean distance of each training observation and the UCL.
n = size(T,1);
D_training = [sqrt(min(sqeucldistm(+T,W.w),[],2)) repmat(W.threshold,n,1)];
%Phase II of the KM-chart:
%Let now R be the (k × p) matrix of quality variables where k is the number
%of observations and p the number of quality variables in the testing phase.
%In Phase II we repeat the same computation as in Phase I but here we use
%test data and compare it with the UCL to detect out-of-control states.
%Compute the Euclidean distance of each test observation.
m= size(R,1);
D_test = [sqrt(min(sqeucldistm(+R,W.w),[],2)) repmat(W.threshold,m,1)];
y1=D_training(:,1); y2=D_training(:,2);x1=(1:n)';
y3=[D_training(:,1); D_test(:,1)];
y4 =[D_training(:,2); D_test(:,2)];x2=(1:n+m)';
%Display the KM-chart for Phase I and II.
figure;
SUBPLOT(2,1,1), plot(x1,y1, '-o', x1,y2, '-'); title('Phase I of the KM-chart')
SUBPLOT(2,1,2), plot(x2,y3, '-o', x2,y4, '-'); title('Phase II of the KM-chart')

```

ALGORITHM 1

The test sample z_j is accepted when its distance is smaller or equal to the radius of the KMDD-based sphere, denoted by R_{KMDD}^2 . This is equivalent to

$$z_j = \begin{cases} \text{in-control observation,} & \text{if } D^2(z_j) \leq R_{\text{KMDD}}^2 \\ \text{out-of-control observation,} & \text{if } D^2(z_j) > R_{\text{KMDD}}^2 \end{cases} \quad (26)$$

where R_{KMDD}^2 is the radius of KMDD-based one-class, representing the UCL for the KM-chart. It is set according to the number of clusters used for the construction of one-class.

4. A Real Industrial Application

To demonstrate the efficacy of the proposed KM-chart, we applied it to the “Cristal Light” cigarettes data set. Actually, “Cristal Light” cigarettes are Tunisian trademark produced by the Kairouan Tobacco Manufacture. The production process of “Cristal Light” cigarettes comprises 12 sequences of operations which are humidification of tobacco leaves, threshing tobacco leaves, strip processing, hashing, drying, expansion of edges, casing, flavoring, introduction of expanded tobacco, confecting ion of cigarettes, packing and boxing, and conditioning. Details about the “Cristal Light” data set can be found in Hajlaoui [26].

The quality of “Cristal Light” cigarettes is defined by five main characteristics which are as follows.

- (1) *The weight of a cigarette*, which is the made up of the tobacco, the filter, and the cigarette paper weights. It varies between 0.965 and 1 gram.
- (2) *The module of a cigarette*, which corresponds to its diameter; it varies from 6.75 to 8.0 millimeters.
- (3) *The humidity rate of tobacco*, which is the proportion of water contained in a cigarette. It is considered acceptable if it varies between 11.5% and 13.5%.
- (4) *Pulling resistance of a cigarette*, which is defined by the difference in pressure between the two extremities of a cigarette when a quantity of air is passed through it. The pulling resistance is considered acceptable when it varies from 100 to 115 CE (colonne d'eau).
- (5) *The folding density*, which corresponds to the volume occupied by the mass of the tobacco inside a cigarette. It is tolerable to belong to $450 \pm 20 \text{ cm}^3$.

The “Cristal Light” data set is composed of 65 observations. The first 60 cigarettes are used to construct OC-charts in phase I. Each cigarette took one minute to be collected. The five remainder cigarettes are used for testing out-of-control states in phase II. For the construction of KM-chart,

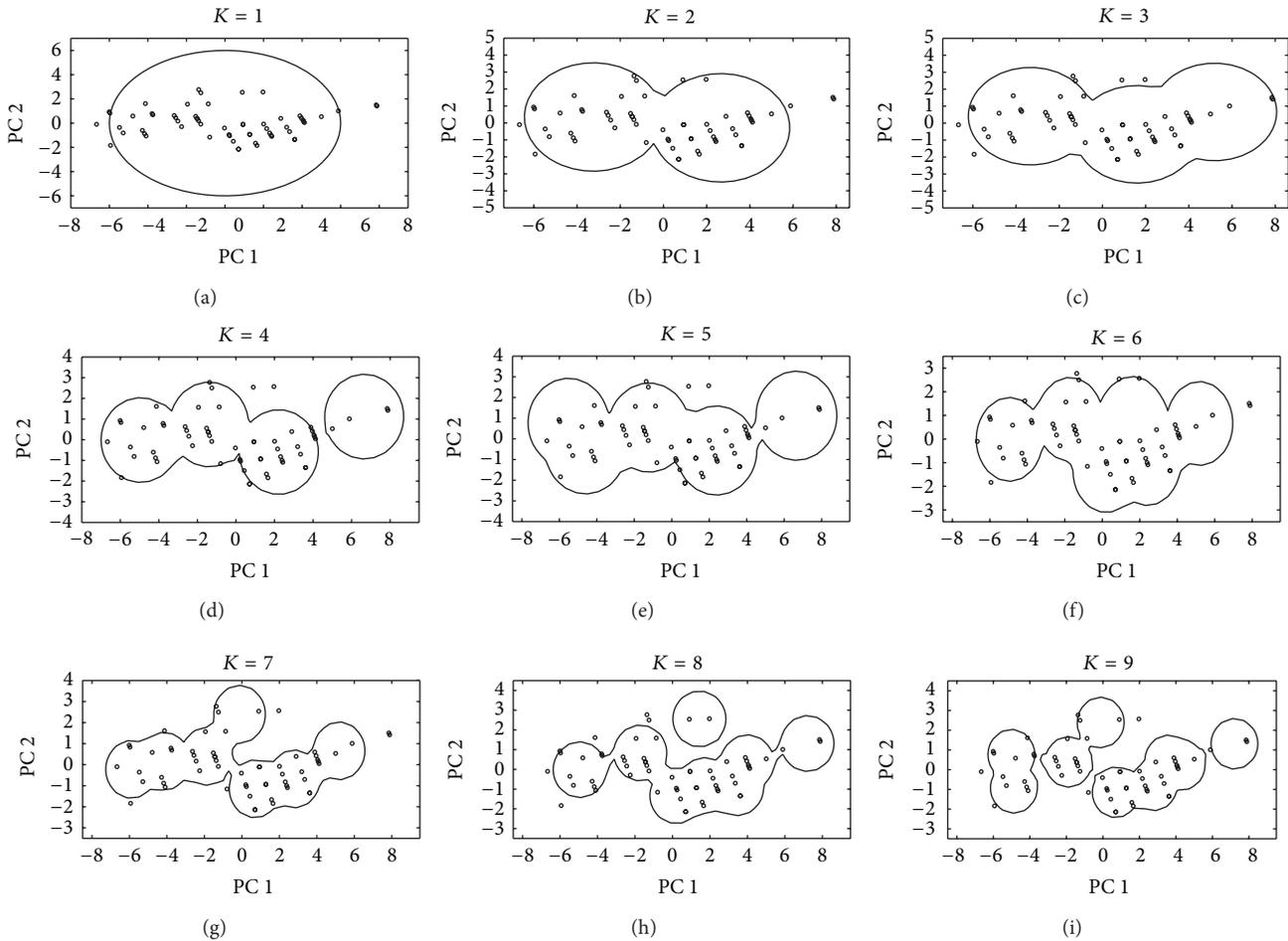


FIGURE 1: Examples of KMDD-based one-classes when varying the number of clusters.

we follow the same methodology of Gani and Limam [25], which consists of three main steps.

Step 1. The data set is analyzed using principal component analysis (PCA) method to obtain independent and identical distributed data, which is a fundamental assumption for one-class classification problem.

Step 2. The principal components (PCs) resulting from Step 1 are used to construct the one-class. In our application, we have three one-class classifiers which are SVDD, KNNDD, and KMDD.

Step 3. The optimal one-class obtained from Step 2 is used to construct OC-charts by computing the charting statistics which are KD for the K -chart, K^2 for the KNN-chart, and D^2 for the KM-chart.

All calculations were carried out with MATLAB software. For the construction of K -chart and KNN-chart, we used the MATLAB codes of Gani and Limam [25]. For the construction of KM-chart, we used the MATLAB code developed by the authors (see Algorithm 1).

After performing PCA, two PCs explaining more than 90% of the variation were retained. Several numbers of clusters were tested for the construction of KMDD-based one-class and to determine the in-control state of the “Cristal Light” process. It is clear from Figure 1 that the number of clusters influences the shape of KMDD-based one-class and plays a pivotal role in determining the trade-off between oversmoothness and undersmoothness of the control boundary. In our application, KMDD-based one-class was constructed with $K = 1$, since the used sample size was not large.

The detection of an abnormal observation in the target class depends on the shape of the established one-class. The KMDD provided a spherical one-class, while SVDD gave a flexible nonspherical one-class due to the use of SVs. It is worth noticing that the shape of SVDD-based one-class depends on the width of the radial basis function while the shape of KNNDD-based one-class is function of the size of the nearest neighbor, denoted by k . Details about the characteristics of SVDD and KNNDD-based one-classes can be found in Gani and Limam [25].

The KM-chart exceeded its control limit of 6.001 at around the 19th, 25th, 28th, 40th, 48th, and 50th cigarettes, as shown in Figure 2. On the other hand, these out-of-control

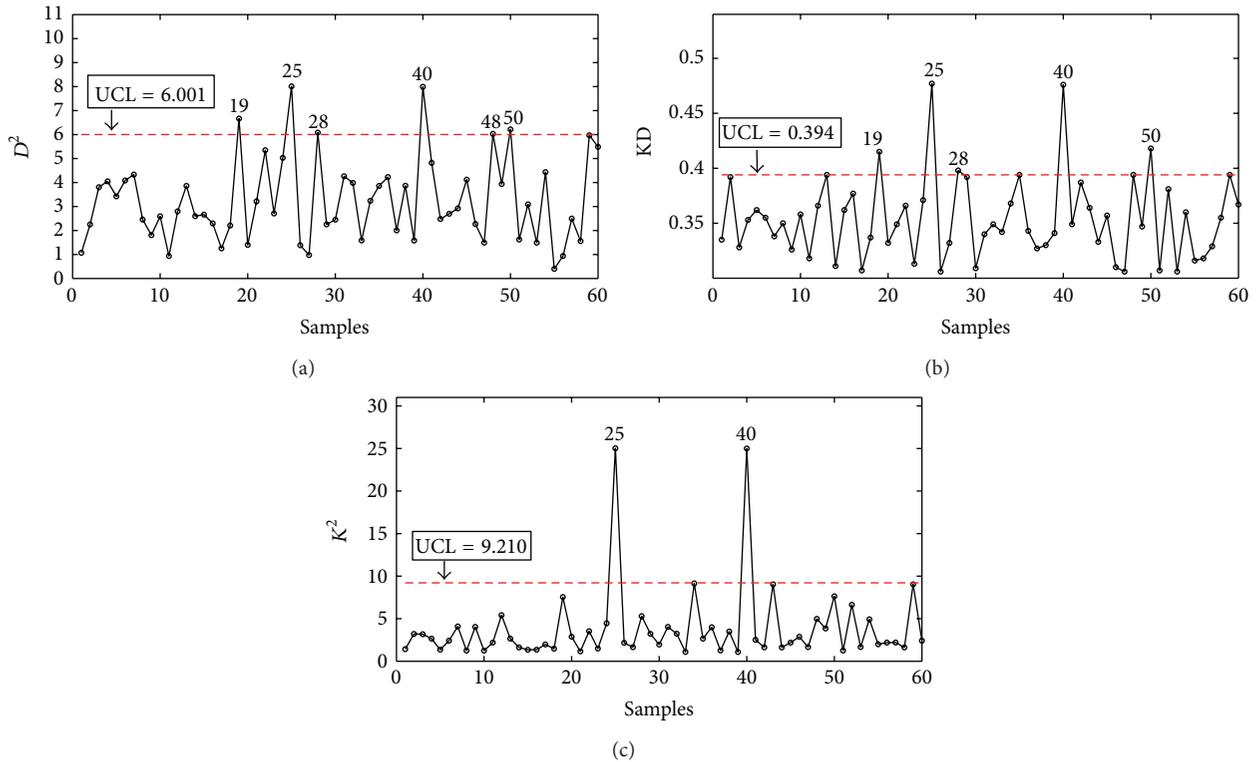


FIGURE 2: Phase I of OC-charts: (a) the KM-chart detected 6 out-of-control cigarettes, (b) the K-chart detected 5 out-of-control cigarettes, and (c) the KNN-chart detected 2 out-of-control cigarettes.

cigarettes have a distance greater than the radius of the established KMDD-based sphere with $K = 1$. For these six abnormal cigarettes, at least one of their five quality characteristics did not respect its tolerance interval, as discussed above. In comparison with the two other control charts, the proposed KM-chart succeeded to detect a new out-of-control observation which is cigarette number 48. Both K-chart and KNN-chart failed to detect this abnormal cigarette. Once these out-of-control observations are removed, no additional outliers were detected, and the in-control process was established.

In phase II, five “Cristal Light” cigarettes were used to detect out-of-control states. The KM-chart triggered an alarm at around the 62nd cigarette and remained below its control limit for the last three cigarettes. Cigarette number 62 was also declared by K-chart as an out-of-control observation, while cigarettes number 62 and 65 were declared by KNN-chart as out-of-control observations. Figure 3 shows the discussed OC-charts for phase II.

5. Performance Comparison

In this section, we study the performance of KM-chart and we compare it with K-chart and KNN-chart. The performance study is based on ARL criterion which is defined as the expected number of samples taken before the shift is detected. It is given by

$$ARL = \frac{1}{p}, \tag{27}$$

where p is the probability of one point plots out-of-control.

A simulation study was conducted to estimate the ARL of OC-charts. In order to be consistent with Gani and Limam [25], we follow their simulation procedure given by the following.

Step 1. Five multivariate normal variables were generated with a mean vector $\mu = (0.986; 7.650; 0.121; 107.183; 451.527)$ and a covariance matrix

$$\Sigma = \begin{bmatrix} 0.0000 & 0.0011 & -0.0000 & -0.0082 & 0.0060 \\ 0.0011 & 0.0889 & -0.0014 & -0.7121 & 0.4635 \\ -0.0000 & -0.0014 & 0.0000 & 0.0120 & -0.0075 \\ -0.0082 & -0.7121 & 0.0120 & 10.7963 & -3.8999 \\ 0.0060 & 0.4635 & -0.0075 & -3.8999 & 2.9820 \end{bmatrix} \tag{28}$$

similar to the mean vector and the covariance matrix of the “Cristal Light” data set used in Section 4. The K-chart, KNN-chart, and KM-chart were designed to achieve an overall in-control ARL of 200. The ARL value was estimated by averaging the run lengths obtained by running 1000 simulated charts.

Step 2. Multivariable shifts, denoted by δ , were introduced in the mean vector according to Table 1. Basically, large values of δ correspond to bigger shifts in the mean. The value $\delta = 0$ is the in-control state.

For detecting small shifts ($\delta = 0.25$), KM-chart performed better than KNN-chart since it gave an ARL = 192.308, while KNN-chart gave an ARL = 200. For the same

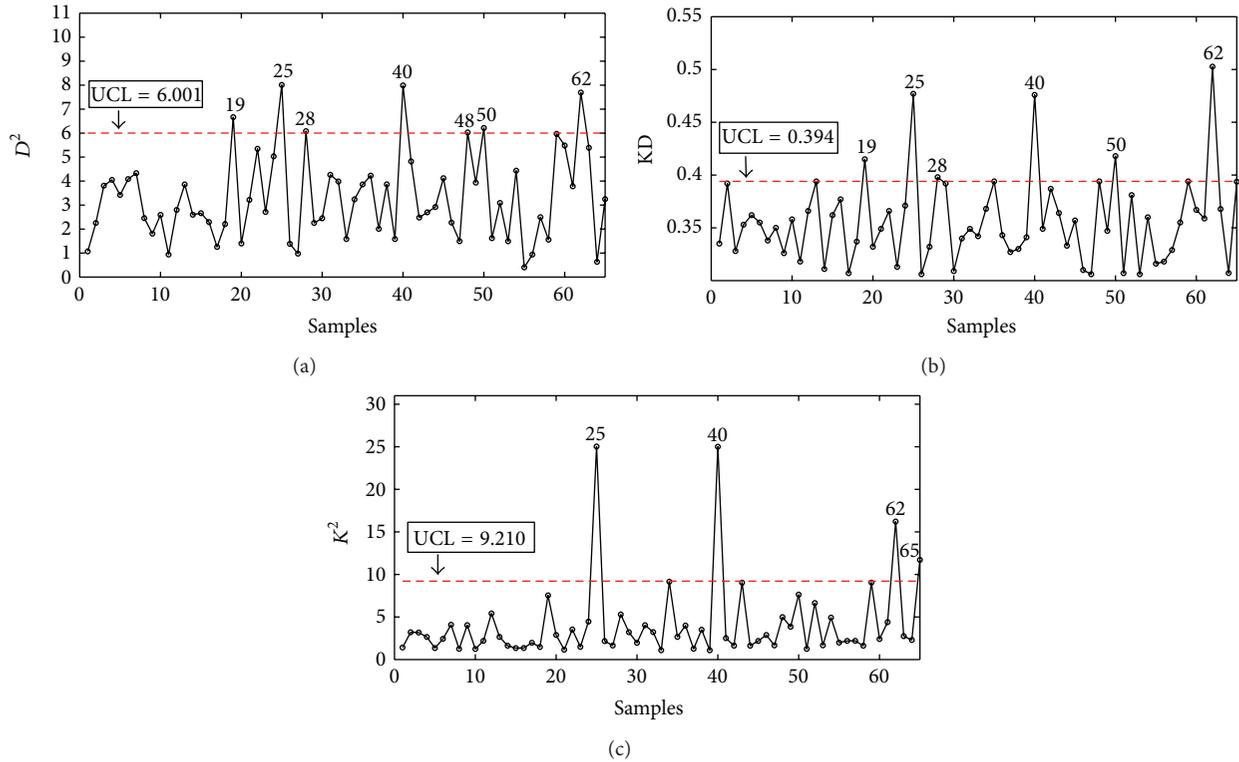


FIGURE 3: Phase II of OC-charts: (a) the KM-chart detected one out-of-control cigarette, (b) the K-chart detected one out-of-control cigarettes, and (c) the KNN-chart detected two out-of-control cigarettes.

TABLE 1: ARL comparison for three OC-charts.

δ	K-chart	KNN-chart	KM-chart
0.00	200.000	200.000	200.000
0.25	100.000	200.000	192.308
0.50	66.667	200.000	185.185
0.75	66.667	100.000	166.667
1.00	50.000	40.000	147.059
1.25	50.000	14.285	117.647
1.50	40.000	9.091	86.956
1.75	28.571	3.703	60.976
2.00	22.222	1.695	39.370
2.25	13.333	1.020	25.707
2.50	6.896	1.000	16.340
2.75	4.167	1.000	9.597
3.00	2.439	1.000	5.540
3.25	1.709	1.000	3.250
3.50	1.197	1.000	1.997
3.75	1.036	1.000	1.390
4.00	1.000	1.000	1.116
5.00	1.000	1.000	1.000

shift level, K-chart yielded an ARL = 100, which was better than that of KM-chart.

For detecting moderate shifts ($\delta = 1$), KNN-chart behaved better than the other two control charts since it gave

an ARL = 40 against an ARL of 50 and 147.059 of K-chart and KM-chart, respectively.

The difference in sensitivity to shifts in the mean vector between the three OC-charts is due to the difference in the nature of distance used by each control chart. The K-chart uses KD, whereas KM-chart and KNN-chart are based on the Euclidean distance. The advantage of KD in comparison with Euclidean distance lies essentially in the use of the kernel function. The latter is equivalent to the distance between two samples measured in a higher dimensional space. This allows K-chart to easily detect any small shift in the process. In terms of ARL and for small shifts in the mean vector, one can draw the conclusion that our proposed KM-chart is situated between KNN-chart and K-chart (KNN-chart < KM-chart < K-chart). Broadly speaking, each OC-chart has its advantages and disadvantages. For example, K-chart performs better than KM-chart and KNN-chart in quickly detecting changes in the process, while the computational cost of KM-chart and KNN-chart is lower than that of K-chart. Table 2 summarizes the characteristics of each OC-chart.

6. Conclusion

In this paper, we have developed a new OC-chart using KMDD algorithm, called KM-chart. Construction methodology of KM-chart is demonstrated through a real industrial application. Performance analysis of KM-chart in phase I and II showed that our proposed control chart is a competitive

TABLE 2: Characteristics of three OC-charts.

	K-chart	KNN-chart	KM-chart
Charting statistic	KD	K^2	D^2
One-class classifier	SVDD	KNNDD	KMDD
Parameters	RBF's width σ	The size of nearest neighbor k	The number of clusters K
Distance type	Kernel distance	Euclidean distance	Euclidean distance
Center	Kernel center	KNN center	Cluster center
Suitable cases	Known and unknown distribution	Known and unknown distribution	Known and unknown distribution

SPC tool. In phase I, the proposed KM-chart detected a new abnormal observation which is cigarette number 48. Both K-chart and KNN-chart failed to detect this abnormal cigarette. Based on the ARL criterion, our proposed control chart outperformed KNN-chart in detecting small shifts in the mean vector.

The proposed KM-chart can be extended to monitor nonlinear processes by using the global kernel K -means algorithm instead of using the standard K -means algorithm. The global kernel K -means has the advantage to identify nonlinearly separable clusters and therefore allows KM-chart to monitor sophisticated manufacturing processes.

Appendix

The MATLAB Code for KM-Chart

The MATLAB code for the KM-chart requires the PRtools toolbox available at <http://www.prtools.org> and the dd_tools toolbox available at http://prlab.tudelft.nl/david-tax/dd_tools.html.

For more details see Algorithm 1.

Conflict of Interests

The authors declare that they do not have a direct financial relation with the software mentioned in this paper and no competing interests.

Acknowledgment

The authors express their appreciation to LARODEC of ISG, University of Tunis, for supporting this paper.

References

- [1] D. F. Cook, "Using basis function networks to recognize shifts in correlated process parameters," *IIE Transactions*, vol. 30, no. 3, pp. 227–234, 1998.
- [2] R. B. Chinnam, "Support vector machines for recognizing shifts in correlated and other manufacturing processes," *International Journal of Production Research*, vol. 40, no. 17, pp. 4449–4466, 2002.
- [3] W. Hwang, G. Runger, and E. Tuv, "Multivariate statistical process control with artificial contrasts," *IIE Transactions*, vol. 39, no. 6, pp. 659–669, 2007.
- [4] J. Hu, G. Runger, and E. Tuv, "Tuned artificial contrasts to detect signals," *International Journal of Production Research*, vol. 45, no. 23, pp. 5527–5534, 2007.
- [5] W. Gani, H. Taleb, and M. Limam, "Support vector regression based residual control charts," *Journal of Applied Statistics*, vol. 37, no. 2, pp. 309–324, 2010.
- [6] S. B. Kim, W. Jitpitaklert, S.-K. Park, and S.-J. Hwang, "Data mining model-based control charts for multivariate and autocorrelated processes," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2073–2081, 2012.
- [7] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *Journal of Process Control*, vol. 19, no. 10, pp. 1627–1639, 2009.
- [8] S. B. Kim, W. Jitpitaklert, and T. Sukchotrat, "One-class classification-based control charts for monitoring autocorrelated multivariate processes," *Communications in Statistics: Simulation and Computation*, vol. 39, no. 3, pp. 461–474, 2010.
- [9] S. Kittiwachana, D. L. S. Ferreira, G. R. Lloyd et al., "One class classifiers for process monitoring illustrated by the application to online HPLC of a continuous process," *Journal of Chemometrics*, vol. 24, no. 3-4, pp. 96–110, 2010.
- [10] S. B. Kim, T. Sukchotrat, and S.-K. Park, "A nonparametric fault isolation approach through one-class classification algorithms," *IIE Transactions*, vol. 43, no. 7, pp. 505–517, 2011.
- [11] R. G. Brereton, "One-class classifiers," *Journal of Chemometrics*, vol. 25, no. 5, pp. 225–246, 2011.
- [12] T. Sukchotrat, S. B. Kim, K.-L. Tsui, and V. C. P. Chen, "Integration of classification algorithms and control chart techniques for monitoring multivariate processes," *Journal of Statistical Computation and Simulation*, vol. 81, no. 12, pp. 1897–1911, 2011.
- [13] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [14] D. M. J. Tax, *One-class classification: concept-learning in the absence of counter-examples [Ph.D. thesis]*, Delft University of Technology, Stevinweg, The Netherlands, 2001.
- [15] R. Sun and F. Tsung, "A kernel-distance-based multivariate control chart using support vector methods," *International Journal of Production Research*, vol. 41, no. 13, pp. 2975–2989, 2003.
- [16] S. Kumar, A. K. Choudhary, M. Kumar, R. Shankar, and M. K. Tiwari, "Kernel distance-based robust support vector methods and its application in developing a robust K-chart," *International Journal of Production Research*, vol. 44, no. 1, pp. 77–96, 2006.
- [17] F. Camci, R. B. Chinnam, and R. D. Ellis, "Robust kernel distance multivariate control chart using support vector principles," *International Journal of Production Research*, vol. 46, no. 18, pp. 5075–5095, 2008.

- [18] W. Gani, H. Taleb, and M. Limam, "An assessment of the kernel-distance-based multivariate control chart through an industrial application," *Quality and Reliability Engineering International*, vol. 27, no. 4, pp. 391–401, 2011.
- [19] W. Gani and M. Limam, "On the use of the K-chart for phase II monitoring of simple linear profiles," *Journal of Quality and Reliability Engineering*, vol. 2013, Article ID 705450, 8 pages, 2013.
- [20] S. He and C. Zhang, "Support vector data description based multivariate cumulative sum control chart," *Advanced Materials Research*, vol. 314–316, pp. 2482–2485, 2011.
- [21] Z. Ge, F. Gao, and Z. Song, "Batch process monitoring based on support vector data description method," *Journal of Process Control*, vol. 21, no. 6, pp. 949–959, 2011.
- [22] X. Liu, K. Li, M. McAfee, and G. W. Irwin, "Improved nonlinear PCA for process monitoring using support vector data description," *Journal of Process Control*, vol. 21, no. 9, pp. 1306–1317, 2011.
- [23] H.-W. Cho, M. K. Jeong, and Y. Kwon, "Support vector data description for calibration monitoring of remotely located microrobotic system," *Journal of Manufacturing Systems*, vol. 25, no. 3, pp. 196–208, 2006.
- [24] T. Sukchotrat, S. B. Kim, and F. Tsung, "One-class classification-based control charts for multivariate process monitoring," *IIE Transactions*, vol. 42, no. 2, pp. 107–120, 2010.
- [25] W. Gani and M. Limam, "Performance evaluation of one-class classification-based control charts through an industrial application," *Quality and Reliability Engineering International*, vol. 29, no. 6, pp. 841–854, 2013.
- [26] M. Hajlaoui, "On the charting procedures: T^2 chart and DD-diagram," *International Journal of Quality, Statistics, and Reliability*, vol. 2011, Article ID 830764, 8 pages, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

