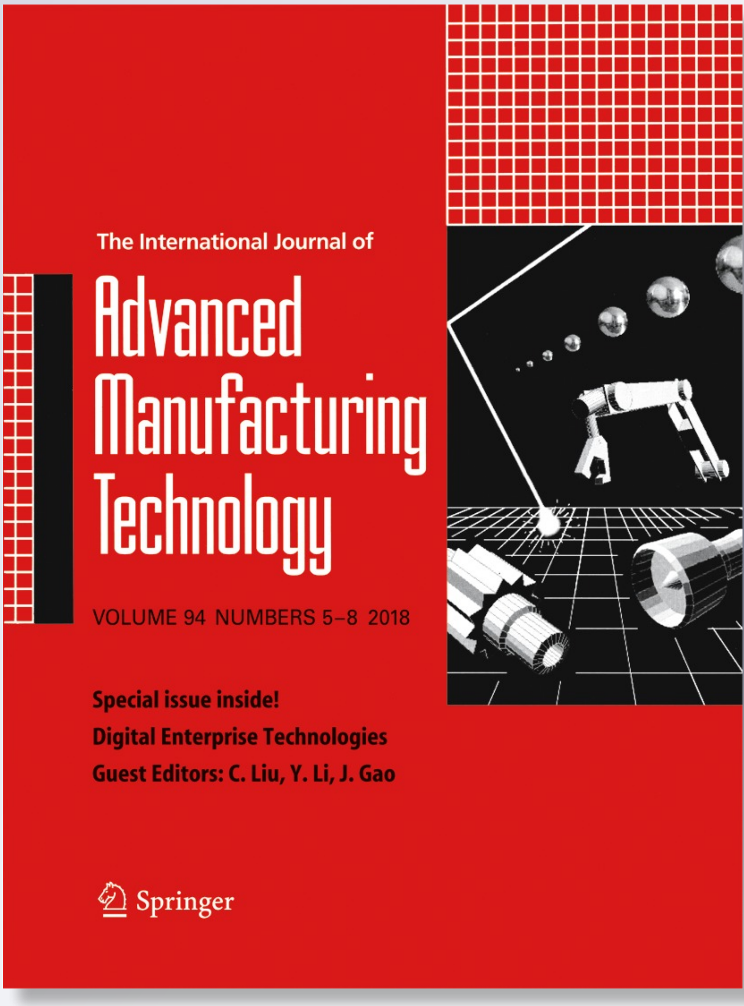
Fault detection and optimal feature selection in automobile motor-head machining process

Edgar O. Reséndiz-Flores, Jesús A. Navarro-Acosta, Cecilia G. Mota-Gutiérrez & Yadira I. Reyes-Carlos

The International Journal of Advanced Manufacturing Technology

ISSN 0268-3768
Volume 94
Combined 5-8

Int J Adv Manuf Technol (2018)
94:2613-2622
DOI 10.1007/s00170-017-1136-x



Springer

Int J Adv Manuf Technol (2018) 94:2613–2622
DOI 10.1007/s00170-017-1136-x

ORIGINAL ARTICLE

Fault detection and optimal feature selection in automobile motor-head machining process

Edgar O. Reséndiz-Flores[1] · Jesús A. Navarro-Acosta[2] · Cecilia G. Mota-Gutiérrez[3] · Yadira I. Reyes-Carlos[3]

Received: 11 August 2017 / Accepted: 13 September 2017 / Published online: 16 September 2017

Abstract This work presents a comparison among four methodologies for fault detection and feature selection applied to dimensional quality control in an automobile motor-head machining process. Three of them are based on the Mahalanobis Taguchi System (MTS) while the fourth one uses Support Vector Data Description (SVDD) one-class classification in order to build an hypersphere with the minimum volume with an enclosed boundary containing almost all target objects. Moreover, Gompertz binary particle swarm optimization (GBPSO) algorithm is applied to optimize kernel hyperparameters in SVDD and simultaneously solve the feature selection problem.

✉ Edgar O. Reséndiz-Flores
eresendiz@itsaltillo.edu.mx

Jesús A. Navarro-Acosta
jesus.navarro@comimsa.com

Cecilia G. Mota-Gutiérrez
cgmota@itsaltillo.edu.mx

Yadira I. Reyes-Carlos
yadira.reyes@itsaltillo.edu.mx

1 Division of Postgraduate Studies and Research, The Technological Institute of Saltillo, Blvd. V. Carranza 2400 Col. Tecnológico C.P. 25280, Saltillo, Coahuila, México

2 Corporación Mexicana de Investigación en Materiales S.A. de C.V. (COMIMSA) Ciencia y Tecnología No 790 Col. Saltillo 400, C.P. 25290, Saltillo, Coahuila, México

3 Department of Industrial Engineering, The Technological Institute of Saltillo, Blvd. V. Carranza 2400 Col. Tecnológico C.P. 25280, Saltillo, Coahuila, México

Keywords Binary particle swarm optimization · Mahalanobis-Taguchi system · Support vector data description · Fault detection · Feature selection

1 Introduction

The use and development of efficient strategies for dimensional reduction on the characteristics for processes quality in order to gain a substantial improvement at the moment of fault detection is crucial. In this aspect, one of the methodologies extensively used is the Mahalanobis-Taguchi System (MTS). MTS is a method proposed by Genichi Taguchi and it does not require statistical assumptions [4, 29]. MTS is proposed as a method of predicting and diagnosing system performance using multivariate data in order to make quantitative decisions with the construction of a multivariate measurement scale using an analytical method. The original version of the Mahalanobis-Taguchi System using orthogonal arrays (OA's) had a wide range of applications since its inception but it has not been without criticism [29]. One of the aspects been criticized is the use of OAs and Signal to Noise Ratio (SNR) for optimal dimensional reduction in order to perform an effective diagnosis. This has resulted in research to implement new strategies to achieve optimal reduction variables like a Binary particle swarm optimization [17, 22] and Gompertz binary particle swarm optimization algorithm [20]. Likewise, machine learning techniques have been applied succesfully in order to perform fault detection in multivariate complex processes which have in common presence of noise and correlated features using large data sets. Based on the statistical learning theory, Support Vector Machines (SVM) is an automatic learning algorithm used for the binary classification and regression of high-dimensional data sets able to deal with highly correlated variables. However, standard binary classifiers show a not satisfactory and poor performance when data sets are unbalanced. Similarly to the hyperplane approach of SVM, Support Vector Data Description (SVDD) estimates a decision hypersphere with the minimum volume to find an enclosed boundary containing only target data in order to perform one class classification dealing in this manner with the unbalanced data sets problem. This paper presents a comparison among four methodologies for fault detection and feature selection applied to dimensional quality control in an automo-bile motor-head machining process. Gompertz binary particle swarm optimization algorithm (GBPSO) is applied to optimize kernel hyperparameters for SVDD and simultaneously solve the feature selection problem. In order to validate the results, also the original MTS, BPSO-MTS, and GBPSO-MTS are presented to compare the performances of the four approaches in terms of the misclassification function. The rest of this paper is organized as follows: Section 2 presents the corresponding background of relevant techniques. The different methodologies are presented in Section 3. Experiments on a real multivariate complex process are shown in Section 4. Finally, in Section 5, conclusions and future works are given.

2 Background

2.1 Mahalanobis-Taguchi System applications

The MTS has been applied in many fields, in particular in the automotive, metal/mechanics and medical branches. The original version of MTS was applied for the first time in a medical diagnosis case and a fire alarm system. The medical diagnosis case is illustrated by taking 17 conditions or features in order to determine a person health. It was found that only 8 variables are considered for diagnostics in the future. The case of fire alarm system was conducted by the Electrical Communication University of Japan. According to the international standards, there are different types of fire situations. A fire alarm system should recognize all these fire situations. Hence, MTS was applied to improve the process monitoring and to improve the functioning of the sensor system by minimizing the false alarms [25]. The application of a MTS-based approach for monitoring and fault identification in chemical processes is presented in [23]. They use stepwise multiple regression analysis instead of orthogonal arrays. MTS is compared to other methods in [28]; the results show that the MTS has better accuracy rate in predicting the discriminant analysis or stepwise discriminant analysis regardless of the complete model or the reduced model. MTS has been used also for multiclass problems, this is the so called Multiclass Mahalanobis-Taguchi System (MMTS), this extension of MTS is developed for simultaneous multiclass classification and feature selection. In MMTS, the multiclass measurement scale is constructed by establishing an individual Mahalanobis space for each class. A real case about gestational diabetes mellitus is studied and the results indicate the practicality of MMTS in real-world applications [24]. There have been applications of the integration of a Neural Network-MTS. In the proposed integrated approach, there exists the following differences with respect to the original MTS: It is not influenced by the multicollinearity problem, the Mahalanobid Distance (MD) computation is an iterative process, and the output signals for normal and abnormal groups have a target value of 0 and 1, respectively. Finally, validation of the separation between output signals for normal and abnormal groups is constantly achieved with testing data sets. It gives flexibility for different applications [18]. In [17], a mathematical model for an optimal variable screening is proposed and solved by binary particle swarm optimization (PSO). In conclusion it was noted that this hybrid MTS-PSO achieves the greatest reduction in multidimensional approach system with respect to SNR. MTS has been integrated to Adaptive Resonance Theory Neural Network for parameter selections in a dynamic product design system (DPDS). The utility of the algorithm is assessed in two dimensions: the MTS shows how individual product parameter dimensions are selected and the ARTN links parameter selection decisions across two different time lines and it can be used to focus on DPDS and to identify product architecture dimensions that are critical for a DPDS [1]. In [21], two alternative optimization methods to the use of OAs in the dimensional reduction variables are compared. The results show that BPSO is faster with respect to speed of convergence to the optimal solution. Regarding dimensional reduction, different scenarios were analyzed. Later on in [22], the optimization problem in reducing variables is solved by applying GBPSO, likewise it was compared against BPSO and NBACO algorithms. The results show that the GBPSO is faster to find the optimal that the BPSO and NBACO. The success rate is 100% for GBPSO in all three cases. A new variant of MTS was developed in [10] proposing a feature selection process that explores a new goodness measure of the model in terms of conditional probability of state system in the subset of variables. The current research presents a methodology for classification based on Mahalanobis Distance (MD) and Association Mining using Rough Sets Theory (RST). In [9], the application of artificial bee colony (ABC) optimization for features selection in MTS is explored. The optimal subset of features is obtained via the stochastic search mechanism of binary ABC. Results are compared with those obtained from two other meta-heuristic techniques, namely genetic algorithm and particle swarm optimization. In [19], Bees Algorithm is deployed replacing the conventional OAs as the optimization technique in MTS. The purpose of BA algorithm in the fusion strategy is to determine the maximum SNR value among the different sets of variables. The optimization problem in reducing variables for the area of laser welding is solved with the application of GBPSO and compared against NBACO algorithm in [20]. The results show that the GBPSO is faster to find the optimum, although both algorithms achieve the optimum value. Moreover, it is the first time that MTS-GBPSO is applied to a laser welding process.

2.2 Support vector data description

Unlike other popular fault detection approaches such as Principal Component Analysis (PCA), Linear Discriminant Analisis (LDA), and the various types of control charts, support vectors methods are able to deal with non linearly separable classification in high dimensional settings in order to perform a robust fault detection procedure since no statistical assumptions are needed. However, in practice, it is common to have few training samples of the faults to be detected (minority class), under this conditions classifiers such as SVM and some others are accurate on the main class but they show very poor accuracy on minority class(es) [7]. Support Vector Data Description (SVDD) estimates a decision hypersphere with the minimum volume to find an enclosed boundary containing only target data in order to perform one class classification and thus deal with the unbalanced data problem. Among the wide range of SVDD applications Guocheng Xie et al. (2013) propose the multiclass hypersphere using support vector data description (HSMC-SVDD) in order to improve the classification accuracy and training speed when the categories have been increased to more than two classes. First, a feature extraction procedure is performed (kurtosis, energy, and correlation). After Nonlinear Principal Component Analysis (KPCA) is implemented to reduce the dimensionality of the data, finally, it develops the hypersphere multiclass support vector data description for diagnosis. Testing is performed on a total of 322 medical images for diagnosis of normal masses, benign masses, or malignant masses [30]. Dong Wang (2013) propose a combination of support vector data description (SVDD) and K-means method for pattern classification. Since K-means representation is too terse and it does not take into account the non-uniform distribution of cluster size. Intuitively, those clusters containing more data are likely to be part of the features with higher influential power compared to the smaller ones. Therefore, SVDD is applied in order to measure the density of each cluster resulted from K-means clustering [27]. Furthermore, the choice of some parameters for SVM-based methods (such as kernel) is of vital relevance for the method performance. Bounsiar et al. (2014) present the research of an appropriate kernel to be implemented in One-Class Classification (OCC) problems. This research is justified because many authors assume that kernels used in standard binary SVM classification are also appropriate to one-class classification. The study presents a breastcCancer classification (benignant or malignant). In this particular case, the radial basis function kernel give satisfactory OCC performance while other kernel functions such as polynomial or sigmoidal kernels are not suitable [3]. Benkedjouh et al (2012) present a method for fault diagnosis and pronosis of bearings based on Principal Component Analysis (PCA) and Support Vector Data Description (SVDD). The purpose of the paper is to transform the monitoring vibration signals into features that can be used to track the health condition of bearings and to estimate their remaining useful life (RUL) [2].

3 Methodologies

3.1 Mahalanobis-Taguchi System

Mahalanobis-Taguchi System (MTS) has gained great interest in the industrial and scientific community since it has shown to be an effective method for fault detection, diagnosis and data classification when applied to multivariate data sets [8, 29]. Usually MTS is implemented in several steps which are described next:

– **Mahalanobis space construction:** data collection corresponding to the multivariate process takes place where the healthy population is defined and the corresponding Mahalanobis distance is computed as follows:

$$MD = \frac{1}{p}(\mathbf{Z}^t\mathbf{C}^{-1}\mathbf{Z})$$

where

$\mathbf{Z}$ = column vector of standarized variables

$\mathbf{C}$ = correlation matrix

p = number of variables

– **Mahalanobis space validation:** the unhealthy population is set and the corresponding MDs are also computed.
– **Identification of useful variables:** In this stage, the estimate S/N analysis for the system using the unhealthy group Mahalanobis distances is done. A robust design in order to reduce the number of important features needed to measure and maintain / improve the S/N analysis is applied where factors are the number of original variables and the number of levels namely are chosen as two, i.e., it might or might not include the variable in the calculation of Mahalanobis distances in

order to improve the efficiency of the system. The positive gains indicate that the variable is useful [8]. Using a larger signal-to-noise ratio is better which is calculated in the following way:

$$S/N^{+} = -10\log\left[\left(\frac{1}{t}\right)\sum_{j=1}^{t}\left(\frac{1}{MD_j}\right)^2\right]$$

where

t = number of replicas

At this point, the best criterion of variables selection must hold in order to avoid a mistaken classification, that is to say, in order to avoid the wrong identification of *healthy* observations in *unhealthy* and vice versa [17].

– **Prediction and future diagnosis:** A new data set is obtained and the corresponding MDs are computed. A control limit CL is defined in terms of the healthy data set. In case the new MDs values overpass the limit CL, the measured information is considered as unhealthy. An appropriate definition of CL is crucial in order to avoid false alarms which imply a great loss in the system [31].

1) Optimum dimensional reduction

The use of orthogonal arrays in the classical form of MTS does not provide an optimal variable selection; therefore, an interesting modification has been proposed where a misclassification problem is considered in order to optimize the real impact of the selected variables and in this way, the important variables in the process are detected in an optimal manner [17]. The optimization problem which is going to substitute the orthogonal arrays approach is stated in terms of the following two misclassification concepts

1. Classifying *healthy* observations as *unhealthy*
2. Classifying *unhealthy* observations as *healthy*

The total misclassification is calculated as follows,

$$w_1 = \frac{c_1}{c_1 + c_2},\; w_2 = \frac{c_2}{c_1 + c_2}$$

Considering these, the main objective function of the total weight is defined as follows:

$$TWFM = w_1\frac{n_1^e}{n_1} + w_2\frac{n_2^e}{n_2}$$

where

n_1^e = number of *healthy* observations classified as *unhealthy*

n_2^e = number of *unhealthy* observations classified as *healthy*

n_1 = number of healthy observations

n_2 = number of unhealthy observations

c_1 and c_2 denote associated costs due to misclassification 1 and 2, respectively. Moreover,

$$n_1^e = \sum_{j=1}^{n_1} MD_j^1 \quad \text{s.t} \quad MD_1^2 \le MD_j^1$$

$$n_2^e = \sum_{j=1}^{n_2} MD_j^2 \quad \text{s.t} \quad MD_{n_1}^2 \le MD_j^2$$

where the following ordering should hold

$$MD_1^1 \le MD_1^1 \le \cdots \le MD_{n_1}^1$$
$$MD_1^2 \le MD_2^2 \le \cdots \le MD_{n_2}^2$$

Regarding the mathematical formulation for selection criteria, a multi-objective program is stated as follows:

$$f_1(\mathbf{x}) = \left(w_1\frac{n_1^e}{n_1} + w_2\frac{n_2^e}{n_2}\right),\; f_2(\mathbf{x}) = \frac{p_{\text{selected}}}{p}$$

Subject to:

$$\sum_{i=1}^{p} x_i \le p, \quad \sum_{i=1}^{p} x_i = p_{\text{selected}}, \quad f_1(\mathbf{x}) \le f_1^{\max}$$

where $\mathbf{x} = (x_1, x_2, ..., x_p)^t$ is a p-dimensional vector and

$$x_i = \begin{cases} 0 & \text{if variable } i \text{ is not selected} \\ 1 & \text{if variable } i \text{ is selected} \end{cases}$$

Finally, the objective function to be minimized reads:

$$f(\mathbf{x}) = \alpha f_1(\mathbf{x}) + \beta f_2(\mathbf{x}) \tag{1}$$

This corresponds to a weighted mixture of binary integer programming where α and β are some predefined weights. With this model, those significant variables in the process are detected in an optimal way in order to monitor and control any multivariate system.

2) Binary PSO (BPSO)

Particle swarm optimization (PSO) is already a well-known algorithm developed by Kennedy and Eberhart, 1995 [11]. This is an stochastic searching technique based on particles population (swarm) that moves within a searching space in a parallel manner in order to find the optimal solution of an objective function.

A particles swarm has to be define in PSO,

$$S = (x_1, x_2, , x_N)$$

where N is the number of particles (possible solutions)

$$x_i = (x_{i1}, x_{i2}, , x_{iP}), i = 1, 2, , N$$

In a similar manner, a velocity vector has to be assigned to each particle which allow the particle to move in the searching space,

$$v_i = (v_{i1}, v_{i2}, , v_{iP}), i = 1, 2, , N$$

The optimization problem must be solved in a binary setting; thus, a binary version of PSO is needed. Kennedy and Eberhart (1997) proposed a binary version of PSO in order to deal with discrete binary variables where the particle velocity is understood in a slightly different manner, i.e., it represents the bit probability of change for each element in the binary particle array [12]. In order to model this probability, the Sigmoid function is used to normalize all values in the range [0,1].

$$sig(v_{ij}(t)) = \frac{1}{1 + e^{-v_{ij}(t)}} \quad (2)$$

The best particle and overall positions are updated by comparing the swarm values. The velocity vector in the binary PSO algorithm is denoted as V_i^c which at the same time it is computed in terms of V_i^1 and V_i^0 which are the probabilities of the particle bit change to 1 and 0, respectively, i.e,

$$V_i^c = \begin{cases} V_i^1 & \text{if } x_{ij} = 0 \\ V_i^0 & \text{if } x_{ij} = 1 \end{cases}$$

P_i^b and P^g positions are updated as in the traditional PSO and the bits carried by them are used to update V_i^0 and V_i^1, with the rules described in the work of Pal and Maiti [17]. Once the new velocity V_i^c is computed, the obtained scalar is mapped to a value $V_{ij} \in [0, 1]$ by a Sigmoid function. Once V_i^c is obtained, the next state of the particle is updated,

$$x_{ik}(t+1) = \begin{cases} 1 - x_{ij}(t) & \text{if } r < sigm(v_{ij}) \\ x_{ij}(t) & \text{if } r > sigm(v_{ij}) \end{cases}$$

In this paper, we refer to this novel binary version of PSO as BPSO.

3) Gompertz BPSO (GBPSO)

The GBPSO algorithm is mainly based on the idea of replacing the sigmoid function as used in the BPSO for generating binary numbers by the Gompertz function given by

$$Gmpz(v_{ij}(t)) = ae^{be^{cv_{ij}(t)}} \quad (3)$$

where a is the upper asymptote usually taken as $a = 1$, while b and c are negative numbers controlling the $v_{ij}(t)$ displacement and the growth rate, respectively. GBPSO has been already studied in [5] where the following forms are proposed to be used:

$$b(t) = -\left(0.2 + (2.0 - 0.2) \cdot \frac{k}{N}\right)$$

$$c(t) = -\left(0.2 + (1.5 - 0.2) \cdot \frac{k}{N}\right)$$

k and N denote the current and maximum number of iterations, respectively. Considering this formulation, the position of the particle is updated using by the following expression,

$$x_{ij}(t+1) = \begin{cases} 1 - x_{ij}(t) & \text{if } r < Gmpz(v_{ij}(t) \\ x_{ij}(t) & \text{another case} \end{cases}$$

3.2 Support vector data description with gompertz BPSO

This strategy carry out reference space optimization and simultaneous feature selection in order to improve fault detection tasks. The goals are the maximization of a model performance and the minimization of the number of used features. The accuracy and stability of the SVDD one-class classifier for fault detection rely on its hyperparameters setting. This study applies an RBF kernel function for the SVDD model to obtain the optimal solution. Thus, parameters C and s must be set properly. Moreover, the variables that contribute for quality defects can be detected by finding the relevant subset of features.

3.2.1 Support vector data description

Consider now that a given data set contains N data objects $\{\mathbf{x}_i\}$, $i = 1, ..., n$. A closed spherical boundary, described by center $\mathbf{a}$ and radius $R > 0$, around the given data must be defined for the target data description. Since the sphere of interest should cover as many data points as possible, the volume of the sphere is minimized over R^2 [13]. Thus, the error function to be minimized reads:

$$\begin{aligned} &\min F(R, \mathbf{a}, \xi_i) = R^2 + C \sum\nolimits_i \xi_i \\ &\text{s.t. } \|\mathbf{x}_i - \mathbf{a}\|^2 \le R^2 + \xi_i, \ \xi_i \ge 0 \ \forall i \end{aligned} \quad (4)$$

where ξ_i denotes slack variables, C controls the trade-off between the volume and the errors. Moreover, if C is too large then we obtain an overfitting model ineffective for the testing phase [6].

Using Lagrange multipliers, the following functional arises:

$$\begin{aligned} L(R, \mathbf{a}, \alpha_i, \gamma_i, \xi_i) = R^2 + C \sum\nolimits_i \xi_i - \sum \alpha_i \\ \times \left\{R^2 + \xi_i - \left(\|\mathbf{x}_i\|^2 - 2\mathbf{a}\cdot\mathbf{x}_i + \|\mathbf{a}\|^2\right)\right\} \\ - \sum\nolimits_i \gamma_i \xi_i \end{aligned} \quad (5)$$

and the corresponding dual form reads:

$$\begin{aligned} &\min \quad \sum\nolimits_i \alpha_i (\mathbf{x}_i, \mathbf{x}_i) - \sum\nolimits_i \sum\nolimits_j \alpha_i \alpha_j \left(\mathbf{x}_i, \mathbf{x}_j\right) \\ &\text{s.t.} \quad 0 \le \alpha \le C \end{aligned} \quad (6)$$

For samples satisfying the equality $\|\mathbf{x}_i - \mathbf{a}\| = R^2 + \xi_i$, the Lagrange multiplier will become unequal to zero ($\alpha_i > 0$).

Those vectors $\mathbf{x}_i$ with $\alpha_i > 0$ are needed to support the hypersphere; thus, they are called the support vectors of the description (SV) [32].

To determine whether a test sample $\mathbf{z}$ is within the hypersphere, its distance to the center of the hypersphere has to be calculated. A test sample $\mathbf{z}$ is accepted when this distance is smaller than the radius [2], i.e.,

$$\|\mathbf{z}-\mathbf{a}\|^2 = (\mathbf{z}\cdot\mathbf{z}) - 2\sum_{i=1}^{l}\alpha_i(\mathbf{z}\cdot\mathbf{x}_i) + \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha j\left(\mathbf{x}_i\cdot\mathbf{x}_j\right) \le R^2 \quad (7)$$

A kernel function is used in place of the inner products in Eqs. 6 and 7 which help to design more flexible methods:

$$\begin{aligned} \min \quad & \sum_i \alpha_i K(\mathbf{x}_i,\mathbf{x}_i) - \sum_i\sum_j \alpha_i\alpha_j K\left(\mathbf{x}_i,\mathbf{x}_j\right) \\ \text{s.t.} \quad & 0 \le \alpha \le C \end{aligned} \quad (8)$$

where $K\left(\mathbf{x}_i,\mathbf{x}_j\right)$ is a kernel function and α_i represents the Lagrange multipliers [15]. One of the most used kernel is the radial basis kernel function (RBF). This kernel is independent of the position of the data set with respect to the origin, it only use the distances between objects [26]. In RBF kernel (9), over/under-fitting depends on values of s [14].

$$K(\mathbf{x}_i-\mathbf{x}_j) = exp\left(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/s^2\right) \quad (9)$$

3.2.2 SVDD with negative examples

When minority classes examples are available (negative examples), they can be used in the training phase to improve the reference space, i.e., these examples must be outside of the hypersphere [16]. Indices i, j are used to identify target vectors while indices p, q are used for negative examples p, q. Furthermore, $y_i = 1$ and $y_p = -1$ denote labels for target and outlier objects, respectively. Relaxing the target as well as the outlier set by allowing some errors and introducing the slack variables ξ_i and ξ_p, then the problem becomes:

$$\begin{aligned} &\min F(R,\mathbf{a},\xi_i,\xi_p) = R^2 + C_1\sum_i \xi_i + C_2\sum_p \xi_p \\ &\text{s.t. } \|\mathbf{x}_i-\mathbf{a}\|^2 \le R^2+\xi_i, \;\; \left\|\mathbf{x}_p-\mathbf{a}\right\|^2 \ge R^2-\xi_p \end{aligned} \quad (10)$$

$\xi_i \ge 0, \xi_p \ge 0 \;\; \forall i, p$. Replacing these constraints in Eq. 10 introducing Lagrange multipliers $\alpha_i, \alpha_p, \gamma_i, \gamma_p$ the following functional is defined:

$$\begin{aligned} & L\left(R,\mathbf{a},\xi_i,\xi_p,\alpha_i,\alpha_p,\gamma_i,\gamma_p\right) \\ &= R^2 + C_1\sum_i \xi_i + C_2\sum_p \xi_p \\ &\quad -\sum_i \gamma_i\xi_i - \sum_p \gamma_p\xi_p - \sum_i \alpha_i\left[R^2+\xi_i-(\mathbf{x}_i-\mathbf{a})^2\right] \\ &\quad -\sum_p \alpha_p\left[(\mathbf{x}_p-\mathbf{a})^2-R^2+\xi_p\right] \end{aligned} \quad (11)$$

with $\alpha_i \ge 0, \alpha_j \ge 0, \gamma_i \ge 0, \gamma_i \ge 0$.
The corresponding dual form reads:

$$\begin{aligned} \min L &= \sum_i \alpha_i(\mathbf{x}_i\cdot\mathbf{x}_i) - \sum_p \alpha_p\left(\mathbf{x}_p\cdot\mathbf{x}_p\right) \\ &\quad -\sum_i\sum_j \alpha_i\alpha_j\left(\mathbf{x}_i\cdot\mathbf{x}_j\right) + 2\sum_p\sum_j \alpha_p\alpha_j \\ &\quad \times\left(\mathbf{x}_p\cdot\mathbf{x}_j\right) - \sum_p\sum_q \alpha_p\alpha_q\left(\mathbf{x}_p\cdot\mathbf{x}_q\right) \end{aligned} \quad (12)$$

By defining new variables $\alpha_i' = y_i\alpha_i$ where index i refers to target and negative vectors. SVDD with negative examples solves in the same manner as the standard SVDD does.

3.2.3 SVDD-GBPSO for reference space optimization and simultaneous feature selection

Let $\mathbf{x}_f = (x_f^1, x_f^2, ..., x_f^p)^t$ be a p dimensional vector where

$$x_f^i = \begin{cases} 0 & \text{if variable} i \text{ is not selected} \\ 1 & \text{if variable} i \text{ is selected} \end{cases}$$

and $\mathbf{x}_C$ and $\mathbf{x}_s \in \mathbb{R}^3$ are coded vectors for C and s values, respectively. Thus, a GBPSO particle is defined as $\mathbf{x_p} = \mathbf{x}_C \cup \mathbf{x}_s \cup \mathbf{x}_f$, see Fig. 1. For reference space optimization and simultaneous feature selection, the optimization problem (1) becomes:

$$\min f(\mathbf{x_p}) = \alpha f_1(\mathbf{x_p}) + \beta\frac{p_{\text{selected}}}{p} \quad (13)$$

subject to:

$$\sum_{i=1}^{p} x_f^i \le p, \quad \sum_{i=1}^{p} x_f^i = p_{\text{selected}}, \quad f_1(\mathbf{x_p}) \le f_1^{\max}$$

where $f_1(\mathbf{x_p})$ is the SVDD performance with selected hyperparameters and optimal subset features. Figure 2 shows how SVDD-GBPSO solves reference space optimization and simultaneous feature selection problem, this approach was taken from [2].

4 Case of study: machining process of the motor-head

The methodologies described in the previous section have been applied to dimensional quality control on an automobile motor-head machining process where 12 characteristics are taken into account. Due to the number of involved features in the process the original MTS can be applied as well as the hybrid methodologies described above. In order

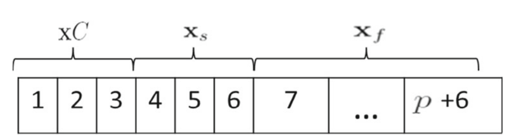


Fig. 1 SVDD-GBPSO Particle

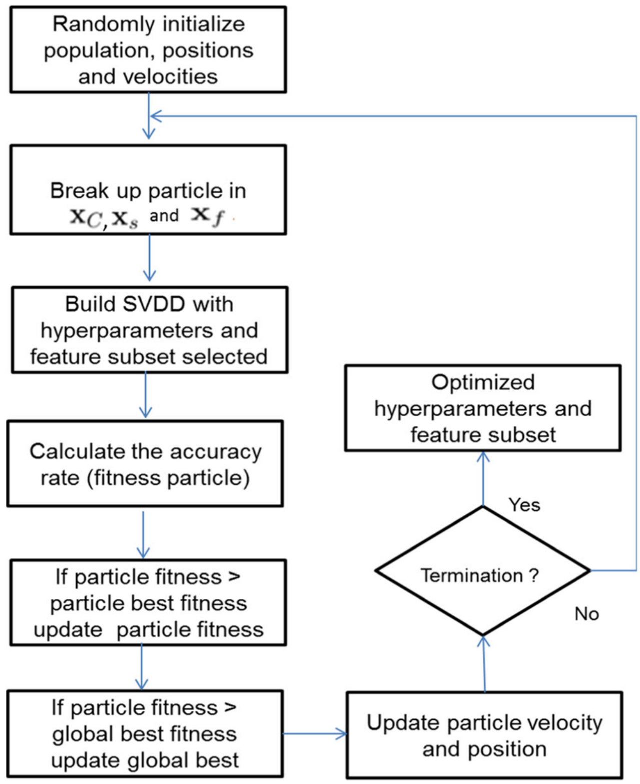


Fig. 2 SVDD-GBPSO diagram

to build up the Mahalanobis space, 189 healthy and 16 unhealthy observations were collected. Table 1 presents the considered variables and their corresponding description.

Table 1 Variables description

Variable	Description
x_1	187-VERTICAL 6MM HOLE 506R INT SIDE
x_2	188-TRANSVERSAL 6MM HOLE 506R INTSIDE
x_3	191-VERTICAL 6MM HOLE 507R EXH SIDE
x_4	192-TRANSVERSAL 6MM HOLE 507R EXH SIDE
x_5	229-LONG.HOLE 124L CAM.L.ADM. REAR 7MM HOLE
x_6	230-TRANS .HOLE 124L CAM.L.ADM. REAR 7MM HOLE
x_7	205-LONG.HOLE 124L CAM.L.ADM. REAR 4MM HOLE
x_8	206-TRANS .HOLE 124L CAM.L.ADM. REAR 4MM HOLE
x_9	235-LONG.HOLE 123L CAM.L.ADM.MIDDLE 7MM HOLE
x_{10}	236-TRANS .HOLE 123L CAM.L.ADM. MIDDLE 7MM HOLE
x_{11}	213-LONG.HOLE 123L CAM.L.ADM.MIDDLE 4MM HOLE
x_{12}	214-TRANS .HOLE 123L CAM.L.ADM. MIDDLE 4MM HOLE

The combinatorial problem for variable screening was solved with swarm intelligence based optimzation algorithms MTS-BPSO, MTS-GBPSO, SVDD-GBPSO, and MTS with orthogonal arrays. The obtained results from phases 1 and 2 are presented below in Fig. 3 which shows that the distances of the “unhealthy” group (red) are greater than the distances of the “healthy” group (blue), validating in this way our Mahalanobis space in a correct manner. Moreover, Table 2 depicts the average of the Mahalanobis distances corresponding to both groups as well as the ranges. An orthogonal array L_{16} has been used in order to resolve the variable/characteristics screening problem where the variables were assigned consecutively in the first 12 columns of the orthogonal array considering the S/N ratios in order to identify which variables have significant impact on process. Table 3 shows the obtained detected variables, in particular x_2, x_5, x_8, which present a positive gain.

We have consider three main configurations in the objective function definition: (1) $\omega_1 = \omega_2 = 0.5$, (2) $\omega_1 = 0.95, \omega_2 = 0.05$, and (3) $\omega_1 = 0.05, \omega_2 = 0.95$ which represent the percentage of importance corresponding to each misclassification function. The weights of the fitness function are fixed to $\alpha = \beta = 0.5$. The number of particles in the swarm for the considered optimization algorithms is 50 and the maximum number of iterations is set to 80.

The objective function optimal values reached after the application of MTS-BPSO, MTS-GBPSO, MTS-Original, and SVDD-GBPSO is reported in Table 4. The selected hyperparameters by SVDD-GBPSO are reported in Table 5.

The optimization performance of the algorithms MTS-BPSO, MTS-GBPSO, and SVDD-GBPSO for the considered three scenarios is depicted in Figs. 4, 5, and 6.

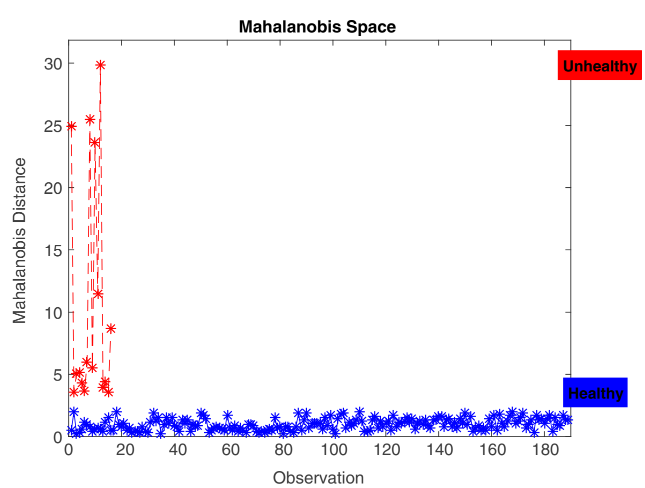


Fig. 3 Mahalanobis space

Table 2 Mahalanobis space

Observations	Average	Range	
		Min	Max
Unhealthy	10.5846	3.5517	29,8431
Healthy	0.9947	0.1869	1.9990

Table 3 Signal-to-noise/ratio

Level	x_1	x_2	x_3	x_4	x_5	x_6
1	1.862	4.329	0.598	2.878	4.129	2.74
0	6.05	3.584	7.314	5.034	3.783	5.172
Gain	−4.188	0.745	−6.716	−2.156	0.346	−2.432
Level	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	3.477	4.782	3.439	2.913	3.606	3.037
0	4.436	3.13	4.473	4.999	4.306	4.875
Gain	−0.959	1.652	−1.034	−2.086	−0.7	−1.838

Table 4 Results

Scenario	Optimal	Variables	Success
$\omega_1 = 0.5$ $\omega_2 = 0.5$			
MTS-Original	0.5763	x_2, x_5, x_8	–
MTS-BPSO	0.31391	$(x_1, x_3, x_5, x_6$	55%
MTS-GBPSO	0.31391	$x_{10}, x_{11}, x_{12})$	41%
SVDD-GBPSO	0.1327	x_1, x_{12}	100%
$\omega_1 = 0.05$ $\omega_2 = 0.95$			
MTS-Original	0.5606	x_2, x_5, x_8	–
MTS-BPSO	0.31391	$(x_1, x_3, x_5, x_6$	56%
MTS-GBPSO	0.31391	$x_{10}, x_{11}, x_{12})$	54%
SVDD-GBPSO	0.0886	x_3, x_{12}	100%
$\omega_1 = 0.95$ $\omega_2 = 0.05$			
MTS-Original	0.5920	x_2, x_5, x_8	–
MTS-BPSO	0.27539	$(x_3$	56%
MTS-GBPSO	0.27539	$x_{10}, x_{12})$	100%
SVDD-GBPSO	0.0812	x_{12}	100%

Table 5 Selected hyperparameters

Scenario	C	s
$\omega_1 = 0.5$ $\omega_2 = 0.5$	1	200
$\omega_1 = 0.05$ $\omega_2 = 0.95$	0.0001	1
$\omega_1 = 0.05$ $\omega_2 = 0.95$	0.0001	1

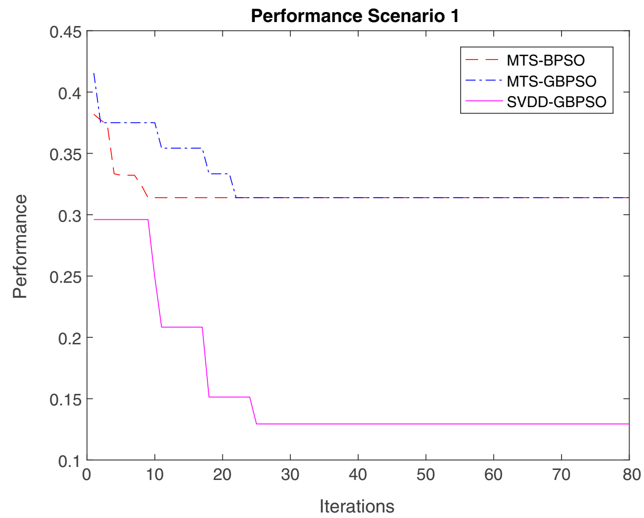


Fig. 4 Scenario $\omega_1 = 0.5$ $\omega_2 = 0.5$

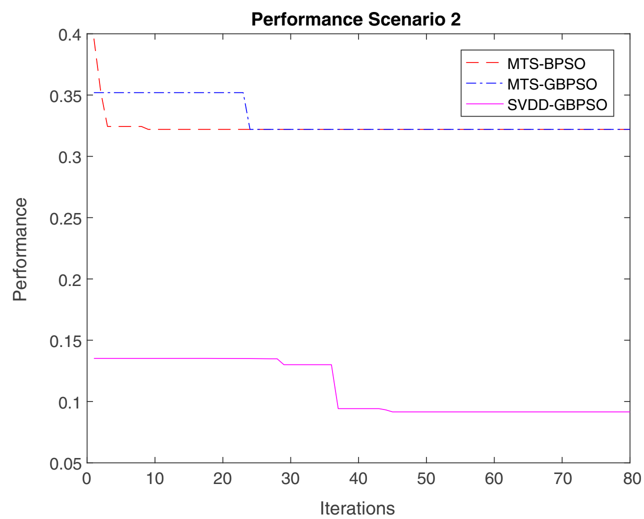


Fig. 5 Scenario $\omega_1 = 0.05$ $\omega_2 = 0.95$

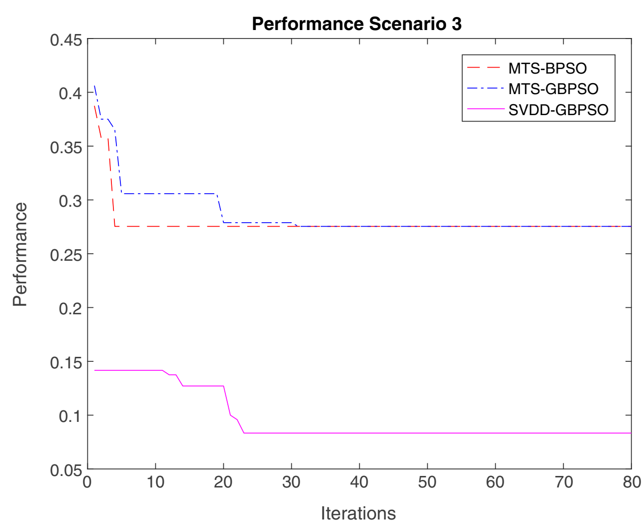


Fig. 6 Scenario $\omega_1 = 0.95$ $\omega_2 = 0.05$

5 Conclusions and future work

Four different methodologies for fault detection and feature selection have been applied for dimensional quality control on an automobile motor-head machining process. MTS-Original, MTS-BPSO, and MTS-GBPSO have been successfully implemented in order to perform optimal dimensional reduction for an effective diagnosis. Furthermore, SVDD-GBPSO approach carry out simultaneous reference space optimization and feature selection in order to improve the fault detection task. It was found that for this particular problem, the SVDD-GBPSO methodology achieves a better performance for fault detection than the other three algorithms in terms of misclassification values and number of selected features. The nature of the SVDD algorithm allows the development of efficient models for classification through the training process with optimal hyperparameters and the selected features, i.e, the optimization of the reference space and feature selection procedure. MTS-based methodologies only perform feature selection in an optimal way while the construction of reference space is based on Mahalanobis distance. Thus, the performance of MTS-based methodologies for fault detection depend to a large extent on the data prior classification into healthy and unhealthy classes. On the contrary, the standard support vectors models such as SVDD algorithm are based on Euclidean distance which is often sub-optimal especially in high dimensions learning problems. As future work, we are interested in the incorporation of the Mahalanobis distance into the SVDD model construction in order to obtain a more robust SVDD classifier. Moreover, testing the performance of other recent metaheuristics such as Artificial Bee Colony optimization, Charged System Search, and Intelligent Water Drops in order to solve the involved combinatorial problem is also of our interest.

References

1. Aman H, Mochiduki N, Yamada H (2012) The mahalanobis taguchi system—adaptive resonance theory neural network algorithm for dynamic product designs. J Inf Optim Sci 33(6):623–635
2. Benkedjouh T, Medjaher K, Zerhouni N, Rechak S (2012) Fault prognostic of bearings by using support vector data description. In: IEEE Conference on Prognostics and Health Management (PHM), pp 1–7
3. Bounsiar A, Madden M (2014) Kernels for one-class support vector machines. In: Information Science and Applications (ICISA). IEEE, pp 1–4
4. Cudney E, Jugulum R, Paryani K (2006) Applying the mahalanobis-taguchi system to vehicle handling. Concurr Eng Res Appl 14(4):343–354
5. Deep K, Chauhan P, Pant M (2012) Multi task selection including part mix, tool allocation and process plans in CNC machining centers using new binary PSO. In: IEEE World Congress on Computational Intelligence, pp 10–15
6. Deng N, Tian Y, Zhang C (2013) Support vector machines, optimization based theory, algorithms, and extensions. Chapman, California
7. Ganganwar V (2012) An overview of classification algorithms for imbalanced datasets. In: International Journal of Emerging Technology and Advanced Engineering
8. Ghasemi E, Asghaie A, Cudney E (2015) Mahalanobis Taguchi system: a review. Int J Qual Reliab Manag 32(3):291–307
9. Iquebal AS, Pal A (2015) Artificial bee colony optimisation-based enhanced Mahalanobis Taguchi system for classification. Int J Intell Eng Inform 2:181–194
10. Iquebal AS, Pal A, Ceglarek D, Tiwari MK (2014) Enhancement of mahalanobis–taguchi system via rough sets based feature selection. Experts Syst Appl 41:8003–8015
11. Kennedy J, Eberhart RC (1995) Particle swarm optimization. IEEE Int Conf Neural Netw 4:1942–1948
12. Kennedy J, Eberhart RC (1997) A discrete binary version of the particle swarm algorithm. IEEE Int Conf Syst Man Cybern 5:4104–4108
13. Lei L, Xiao-dan W, Xi L, Ya-fei S (2015) Hierarchical error-correcting output codes based on SVDD. pattern analysis and applications. Springer 19:163–171
14. Lin S, Lee Z, Chen S, Tsen T (2008) Parameter determination of support vector machine and feature selection using simulated annealing approach. Appl Soft Comput 8:1505–1512
15. Liu J, Sun Y (2013) Multivariate statistical process monitoring scheme with PLS and SVDD. In: International Conference on Industrial Engineering and Engineering Management
16. Mu T, Nandi A (2009) Multiclass classification based on extended support vector data description. IEEE Trans Syst Man Cybern 39:1206–1216
17. Pal A, Maiti J (2010) Development of a hybrid methodology for dimensionality reduction in Mahalanobis-Taguchi system using Mahalanobis distance and binary particle swarm optimization. Expert Syst Appl 37(2):1286–1293
18. Rai B (2009) Classification, feature selection and prediction with Neural-network Taguchi System. Int J Ind Syst Eng 4(6):645–664
19. Ramlie F, Jamaludin K, Dolah R (2016) Optimal feature selection of taguchi character recognition in the mahalanobis-taguchi system using bees algorithm. Glob J Pure Appl Math 12(3):2651–2671
20. Reséndiz-Flores EO, López-Quintero ME (2017) Optimal identification of impact variables in a welding process for automobile seats mechanism by MTS-GBPSO approach. Int J Adv Manuf Technol 90(1):437–443
21. Reséndiz-Flores EO, Moncayo-Martínez LA, Solís G (2013) Binary ant colony optimization applied to variable screening in the Mahalanobis-Taguchi system. Expert Syst Appl 40(2):634–637
22. Reséndiz-Flores EO, Rull-Flores CA (2013) Mahalanobis-Taguchi system applied to variable selection in automotive pedals components using gompertz binary particle swarm optimization. Expert Syst Appl 40(7):2361–2365
23. Saraiva P, Faísca N, Costa R, Gonçalves A (2004) Fault identification in chemical processes through a modified Mahalanobis–Taguchi strategy. Comput Aided Chem Eng 18:799–804
24. Su C, Hsiao Y (2007) An evaluation of the robustness of MTS for imbalanced data. IEEE Trans Knowl Data Eng 19(10):1321–1332
25. Taguchi G, Jugulum R (2000) New trends in multivariate diagnosis. Sankhyā: Indian. J Stat Ser B 62(2):233–248
26. Tax D, Duin R (2004) Support vector data description. Mach Learn 54:45–66
27. Wang D, Tan X (2013) Neural information processing. Springer, Berlin

28. Wang H, Chiu C, Su C (2004) Data classification using the Mahalanobis-Taguchi system. J Chin Inst Ind Eng 21(6):606–618
29. Woodall WH, Koudelik R, Tsui KL, Kim SB, Stoumbus ZG, Carvounis CP (2003) A review and analysis of the mahalanobis-taguchi system. american statistical association and the american society for quality. Technometrics 45(1):1–15
30. Xie G, Jiang Y, Chen N (2013) A multi-class support vector data description approach for classification of medical image. In: Ninth International Conference on Computational Intelligence and Security, pp 115–119
31. Yang T, Cheng Y (2010) The use of Mahalanobis-Taguchi System to improve flip-chip bumping height inspection efficiency. Microelectron Reliab 50(3):407–414
32. Yin G, Zhang Y, Li Z, Ren G, Fan H (2014) Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure. Elsevier, Neurocomputing