Feature Selection for Mahalanobis-Taguchi System with Chaotic Quantum Behavior Particle Swarm Optimization

Xiao-hong PENG¹, Rui ZHENG² and Jiu-fu LIU²,*

¹School of Information Engineering, Jiangsu Maritime Vocational Institute Nanjing 211170, China
²College of Automation, Nanjing University of Aeronautics and Astronautics Nanjing 210016, China

*Corresponding author

Keywords: Mahalanobis–Taguchi system, Variable optimization method, Chaotic quantum behavior particle swarm, Optimization.

Abstract. This paper proposes a Mahalanobis-Taguchi system variable optimization method based on chaotic quantum behavior particle swarm, and solve the MTS multi-objective optimization problem by using the chaotic quantum behavior particle swarm optimization algorithm to reduce the feature variables and improve the computational efficiency and accuracy of optimization.

Introduction

The Malalanobis-Taguchi system (MTS) is a new method for pattern recognition and classification, diagnosis and prediction of multi-featured variable data sets [1-5]. In the process of system feature selection optimization, The traditional Mahalanobis-Taguchi system calculates the signal-to-noise ratio to optimize variable combination by the orthogonal table. However, the orthogonal table optimization is inefficient in practical application, and the actual system classification performance after screening is difficult to achieve the expected effect.

Zeng et al [6] proposed the use of robust design principles and fuzzy mathematics to optimize the space of MTS. Iquebala et al.[7] introduced a rough set to establish an optimization model, and used signal-to-noise ratio and orthogonal array methods to optimize the selection of characteristic variables. Niu[8] comprehensively considered the classification accuracy, the large signal-to-noise ratio and the degree of dimensionality reduction, and constructed a multi-objective optimization model. This method can lead to local optimization results.

Reséndiz et al. [9] used the binary ant colony algorithm instead of the orthogonal table screening method in MTS system optimization, and the optimization speed is slow. Pal et al. [10] applied binary particle swarm optimization (BPSO) to the MTS model dimension reduction, and the computational speed was greatly improved, but the optimization results tend to be locally optimal. The particle convergence of the traditional PSO algorithm is realized in the form of orbit. At the same time, the particle velocity is limited, so that the search range is a finite region, and it is difficult to cover the entire solution space. This leads to the PSO algorithm not guaranteeing the convergence to the global with probability 1. The Quantum behaved Particle Swarm Optimization (QPSO) theory introduces particles into quantum space. Particles in quantum bound states can reach arbitrary points in space depending on a certain probability density. The feasible solution space can reach the global and will not diverge to infinity.

Due to the advantages of quantum particle swarm optimization on variable optimization, this paper combines the algorithm with MTS model and uses QPSO to reduce the feature variables.

Mahalanobis-Taguchi System

The Mahalanobis-Taguchi system uses the Mahalanobis distance (MD) [11], signal-to-noise ratio (SNR) and orthogonal table as tools to complete the diagnosis and prediction of the system. The specific steps are as follows:

Step 1: Build the reference space
First, it is necessary to determine normal and abnormal samples from the collected raw sample
data, and select appropriate feature variables. $n$ samples of the normal group is selected, and the number of characteristic variables is $k$. According to the following formula (1), the vector group is normalized. Then the inverse matrix method is used to calculate the Mahalanobis distance of each normal sample according to the following formula (2) to construct measure reference space.

$$z_j = \frac{1}{s_j} (x_j - \bar{x}_j) \quad i=1,2,\ldots,n, \quad j=1,2,\ldots,k$$

(1)

$$MD_i = d_i^2 = \frac{1}{k} \mathbf{Z}_i^T \mathbf{S}^{-1} \mathbf{Z}_i \quad i=1,2,\ldots,n$$

(2)

Here, $\mathbf{Z}_i = [z_{i1}, z_{i2}, \ldots, z_{ik}]$, $i=1,2,\ldots,n$ is the normalized vector of the $i$th sample. $\bar{x}_j$ represents the data mean of all samples of the $j$th dimension variable. $s_j$ represents the data standard deviation of all samples of the $j$th dimension variable. $\mathbf{S}^{-1}$ represents the inverse matrix of the correlation matrix of the sample population.

Step 2: Verify the validity of the reference space

Generally, $n$ the basis of the reference space, the mean, standard deviation and correlation matrix of the normal samples are still used. The Mahalanobis distance is calculated for the training data set. If the MD of the abnormal sample is larger than the MD of the reference, and the data of the two categories can be well distinguished, the reference space is valid.

Step 3: Optimization of the reference space

MTS introduces orthogonal table and signal-to-noise ratio to optimize the feature attributes of the system. The two-level orthogonal table is used to select the appropriate specifications according to the number of features. The experiments are sequentially performed according to the selected variables of each row respectively. Equations (2) and (3) are used to calculate the MD and SNR values of the anomaly samples, respectively.

$$SN_j = -10 \log \left[ \frac{1}{n} \sum_{i=1}^{n} (1 / MD_i^2) \right]$$

(3)

For the variable $X_j$, the SNR average of all the selected variables in the orthogonal table is calculated, denoted as $SN_{j+}$. Then the SNR average value of the variable that is not selected is calculated, as $SN_{j-}$, and $SN_{j+} - SN_{j-}$ is used to represent the gain value of the variable. If the gain value of the variable is positive, it is retained, otherwise it is eliminated.

Step 4: Determine the threshold and predict the classification

The classification threshold $T$ is determined, and the optimized measurement space is used to calculate the MD of each unknown sample. Finally, the unknown sample realizes the diagnosis and prediction of the system according to the threshold. For the two-category problem, the sample is divided into two types: normal class and abnormal class. If the sample is calculated that the MD is smaller than the threshold $T$, it is classified as a normal class, and vice versa.

**Dimensionality Reduction Model**

**Gram-Schmidt Orthogonalization Method for Obtaining Mahalanobis Distance**

The Mahalanobis distance is a measure of the degree of similarity between samples used in MTS [7]. Considering that there will be multiple collinearity between variables, the correlation matrix between samples cannot be inverted. The Mahalanobis-Taguchi-Gram-Schmidt system (MTGS) [8] proposes Gram-Schmidt orthogonality. The Mahalanobis distance is obtained by the Gram-Schmidt Orthogonalization process (GSP). This method can make the properties of the sample completely orthogonal to avoid the influence of multicollinearity on the distance metric.

The Gram-Schmidt orthogonalization process firstly normalizes the initial vector group according
to equation (1), and then computes an orthogonal set of vectors for the linear independent vector group. The calculation process of GSP is as follows.

\[ U_i = Z_i \]
\[ U_2 = Z_2 - \frac{Z_i^T U_1}{U_1^T U_1} U_1 \]
\[ \ldots \]
\[ U_k = Z_k - \frac{Z_i^T U_1}{U_1^T U_1} U_1 - \ldots \frac{Z_i^T U_{k-1}}{U_{k-1}^T U_{k-1}} U_{k-1} \]  \( (4) \)

Where \( k \) is the total number of variables, \( n \) is the total number of samples. \( Z_j = (z_{j1}, z_{j2}, \ldots, z_{jn})^T \), \( j = 1,2,\ldots,k \) is the jth dimensional normalized vector of the \( n \) samples. \( U_k = (u_{ik}, u_{2k}, \ldots, u_{nk})^T \) is the mutual vertical vector of the kth dimension with the same linear span.

Then the Mahalanobis distance of the ith sample is given by:

\[ MD_i = \frac{1}{k} \left( \frac{u_{i1}^2}{s_1^2} + \frac{u_{i2}^2}{s_2^2} + \ldots + \frac{u_{in}^2}{s_n^2} \right) \]  \( (5) \)

Where \( u_{ik} \) is the element of the orthogonal vector \( U_k \), \( s_k \) is the standard deviation of each element in the vector \( U_k \).

**Multi-objective Optimization Model**

MTS is usually used for the problem of two classifications. The main goal of MTS in variable optimization reduction is to use a combination of variables that are de-redundant to make the error probability of the classification as low as possible.

**Goal 1:** The concept of misclassification rate is defined as the criterion for variable screening. The severity of the consequences of two different types of misclassifications may vary depending on the system being considered. The measurement of misclassification rate is calculated by the weighted sum of different objective functions. The combination of variables to be optimized can be defined as \( X = (x_1, x_2, \ldots, x_k) \).

\[ x_i = \begin{cases} 1, \text{ If the ith variable is selected} \\ 0, \text{ If the ith variable is not selected} \end{cases} \]

The expression for the misclassification rate is as follows:

\[ \begin{cases} f_1(X) = \frac{n'_1}{n_1} \\ f_2(X) = \frac{n'_2}{n_2} \end{cases} \]  \( (6) \)

Here, \( n'_1 \) indicates the number of normal observations misclassified as abnormal. \( n'_2 \) indicates the number of abnormal observations misclassified as normal. \( n_1 \) represents the total number of normal group samples. \( n_2 \) represents the total number of abnormal group samples.

The values of \( n'_1 \) and \( n'_2 \) are obtained by the following functions Get_n'_1( ) and Get_n'_2( ):

\[ \text{Get}_n'_1( MD^1 ) \]
\[ n'_1 = 0; \]

for \( j = 1 \) to \( n_1 \) (the number of normal samples) do

If \( MD_j > T \)  \( n'_1 = n'_1 + 1; \)
Endif
Return \( n^1_i \)

Get\(_{n^2_j} (MD^2)\)
\( n^2_i = 0; \)

for \( j=1 \) to \( n^2 \) (the number of abnormal samples) do
If \( MD^2_j \leq T \) \( n^2_i = n^2_i + 1; \)
Endif
Return \( n^1_i \)

The function input parameters \( MD^1 = \{MD^1_1, MD^1_2, \ldots, MD^1_n\} \) and \( MD^2 = \{MD^2_1, MD^2_2, \ldots, MD^2_n\} \) are the Mahalanobis distance values of the normal group and the abnormal group calculated using the optimized variables, respectively. Here, the ROC curve method is used to determine the optimal threshold point of the system.

Goal 2: Minimize the number of characteristic variables of the system by dimension reduction, and see the following expression:

\[
f_3 = \frac{p_{\text{selected}}}{p}
\]

The multi-objective optimization model is listed below:

\[
\min F(X) = \omega_1 \frac{n^1_i}{n_1} + \omega_2 \frac{n^2_i}{n_2} + \omega_3 \frac{p_{\text{selected}}}{p}
\]

s.t.

\[
\sum_{i=1}^{p} x_i \leq p
\]

\[
\sum_{i=1}^{p} x_i = p_{\text{selected}}
\]

\[
f_1(X) \leq f_1^{\max}, \ f_2(X) \leq f_2^{\max}
\]

In this paper, the above expression (8-11) is used to reduce the dimension of MTS variables.

**Optimization Solution**

In this paper, an improved binary-coded chaotic quantum particle swarm optimization algorithm is introduced for the optimal solution of the model.

**Chaotic Quantum particle Swarm Optimization (CQPSO)**

In order to enhance the randomness of particle swarms and jump out of the local search region, the algorithm proposes the concept of local attractors. which can be seen in the following formula:

\[
p_i(t) = \phi p_{i_{\text{best}}}(t) + (1-\phi)G_{j_{\text{best}}}(t)
\]

Where \( p_i(t) \) is the attractor of the \( j \)th dimension of the \( i \)th particle, and \( p_{i_{\text{best}}}(t) \) represents the local optimal position of the \( j \)th dimension of the \( i \)th particle. \( \phi \) is the random number uniformly distributed over the interval \( \{0, 1\} \).

The chaotic mapping relationship is determined according to the distance between the position of the \( i \)-th particle and \( p_i(t) \). Then the chaotic search range of the next generation particle is calculated according to the Monte Carlo concept:
\[ x_{i \text{min}} = \begin{cases} 
\frac{p_i(t) - p_i(t)\cdot x_i(t)}{u}, & \text{if } p_i(t) \neq x_i(t) \\
0, & \text{otherwise}
\end{cases} \tag{13} \]

\[ x_{i \text{max}} = \begin{cases} 
\frac{p_i(t) + p_i(t)\cdot x_i(t)}{v}, & \text{if } p_i(t) \neq x_i(t) \\
0, & \text{otherwise}
\end{cases} \tag{14} \]

Where, \( u, v \in (0,1) \).

Then, the particles are normalized according to Equation (15), and \( Q_{i,1} \), the initial value of the chaotic sequence is calculated.

\[ Q_{i,1} = \frac{x_i(t) - x_{i \text{min}}}{x_{i \text{max}} - x_{i \text{min}}} \tag{15} \]

\( Q = (Q_{1,1}, Q_{2,1}, ..., Q_{m,1}) \) which is the chaotic sequence, is calculated according to the chaotic map structure selected by the algorithm, and \( m \) is the length of the chaotic sequence. The chaotic sequence is inversely transformed into the original space according to Equation (16).

\[ F(x_i, r(t)) = x_{i \text{min}} + Q_{i,1}(x_{i \text{max}} - x_{i \text{min}}) \tag{16} \]

Finally, the particles with the best evaluation results in the sequence are updated to the next generation particles. The iterative formula of the CQPSO algorithm is shown in (17):

\[ x_i(t+1) = \arg \min_{1 \leq r \leq m} F(x_i, r(t)) \tag{17} \]

**Particle Coding Principle**

The CQPSO particle is a continuous string of values, it cannot be used for the calculation of the particle's objective function[12]. Therefore, it is necessary to convert the position value of the particle in binary form. The idea of the conversion is to map the particle position to the sigmoid function. According to the position value of the particle, the probability value in the interval \( \{0,1\} \) can be obtained. The particle selects whether to update the position according to the size of the probability value. The formula is based on the following:

\[ x_{i}(t + 1) = \begin{cases} 
0, & \text{if } \text{sig}(x_{i}(t)) < \text{rand()}
\\
1, & \text{if } \text{sig}(x_{i}(t)) \geq \text{rand()}
\end{cases} \tag{18} \]

Here \( \text{rand()} \) is a random number in the interval \( \{0,1\} \), and the sigmoid function is expressed as follows:

\[ \text{sig}(x) = \frac{1}{1 + e^{-x}} \]

The sequence of variables after conversion to a binary string by successive values can be used to calculate the value of the objective function and perform subsequent valid updates.

**Variable Reduction Algorithm Flow of Quantum Behavior Binary Particle Swarm**

This section proposes an improved variable reduction algorithm for chaotic quantum behaved particle swarm optimization. The input of the algorithm is the sequence of characteristic variables, the size of the population, the dimension of the particle, and the upper and lower bounds of the particle. The output is the global optimal fitness value of the particle, and the global optimal particle sequence.

**Algorithm:** Chaotic quantum behavior particle swarm optimization

**Input:** sequence of characteristic variables \( X = (x_1, x_2, ..., x_i) \), population size \( N \), dimension of particles \( \dim \), maximum value of chaotic map sequence \( m \)

**Output:** the global optimal fitness value of the particle \( F_{\text{gbest}} \), the global optimal particle sequence \( X_{\text{gbest}} \)
Randomly initializes the current position of each particle in the particle swarm $x_i$, determines the maximum number of iterations $MaxIter$.

The particle position $x_i(t)$ is binary coded according to equation (18), and the fitness function value of each particle is calculated, determines the individual optimal position $p_{best}(t)$ and the global optimal position $G_{best}(t)$.

1. Randomly initializes the current position of each particle in the particle swarm $x_i$, determines the maximum number of iterations $MaxIter$.
2. The particle position $x_i(t)$ is binary coded according to equation (18), and the fitness function value of each particle is calculated, determines the individual optimal position $p_{best}(t)$ and the global optimal position $G_{best}(t)$.
3. for ($i=1;i \leq N;i=i+1$) do
4. Calculate the local attractor $p_{g}(t)$ according to equation (12)
5. Calculate $(x_{i_{min}}, x_{i_{max}})$, the chaotic search range of next generation particles according to equations (13) and (14)
6. Calculate $Q_{i,1}$, the initial value of the chaotic sequence according to equation (15)
7. repeat
8. Calculate $Q_{i,r}$, the Logistic chaotic sequence according to equation $f(Q_{i,r+1})=4Q_{i,r}(1-Q_{i,r})$
9. Mapping the inverse transform chaotic to the original space according to equation (15)
10. Binary encoding the particle position $x_{i}(t)$ according to equation (17)
11. Calculate the fitness function value of the current particle and update the optimal position of the particles $p_{best}(t)$, $G_{best}(t)$
12. Linearly decreasing the length $m$ of the chaotic map sequence
13. Until the maximum number of iterations $MaxIter$ is reached
14. End for
15. Output global optimal fitness value $F_{global}$ and optimized sequence of particles $X_{best}$

Case Study

Experimental Data

This paper conducts a case study on the steel plate fault data set provided by the Semeion Communication Science Research Center. The data set contains 25 fault type labels, which can be seen in Table 1. This is a high-dimensional data set. There may be redundancy of some attributes, and it is necessary to reduce the dimension of system. In this experiment, the category of other defect was used as normal sample and the category of K scratch was used as abnormal sample. 512 normal samples were selected, including 419 sets of training data and 93 sets of test data. Additionally, we selected 295 sets of abnormal samples, including 210 sets of training data and 85 sets of test data.

| Numbers | Attributes | Numbers | Attributes |
|---------|------------|---------|------------|
| $x_1$   | X Minimum  | $x_{14}$ | Empty index|
| $x_2$   | X Maximum  | $x_{15}$ | Square index|
| $x_3$   | Y Minimum  | $x_{16}$ | Outside X index|
| $x_4$   | Y Maximum  | $x_{17}$ | Edges X index|
| $x_5$   | Pixels areas | $x_{18}$ | Edges Y index|
| $x_6$   | X perimeter | $x_{19}$ | Outside global index|
| $x_7$   | Y perimeter | $x_{20}$ | Log of areas|
| $x_8$   | Sum of luminosity | $x_{21}$ | Log X index|
| $x_9$   | Minimum of luminosity | $x_{22}$ | Log Y index|
| $x_{10}$ | Maximum of luminosity | $x_{23}$ | Orientation index|
| $x_{11}$ | Length of conveyer | $x_{24}$ | Luminosity index|
| $x_{12}$ | Steel plate thickness | $x_{25}$ | Sigmoid of areas|
| $x_{13}$ | Edges index |
Data Processing and Analysis

The experiment shows that the normal sample in the reference space has an MD mean of 0.98 and its value is close to 1. The MD value of the abnormal sample is calculated, and its value is significantly larger than the normal sample, so the reference space is confirmed to be valid. Then, using the initial characteristic variable sequence, the MD of the remaining 219 normal samples of the training data set and the MD of the abnormal sample training data set are calculated, which can be seen in the following scatter Fig. 1.

![Figure 1. MD value distribution of normal and abnormal group samples.](image)

According to Fig. 1, it can be seen that the MD value of the normal group is small, and the average value is 2.73. The MD value of the abnormal group fluctuates greatly, and the mean value reaches 24.9. The overall MD has a certain degree of discrimination, but there are still more overlapping parts and further classification is more difficult. The optimal threshold of the system at this time is determined by the ROC curve method to be 2.16, and the prediction accuracy is only 80.23%.

It can be seen that the original variable combination of the sample data set is not ideal for the identification of the system. There are some redundant variables that have a negative effect on the classification result. Therefore, it is necessary to optimize the multi-dimensional attributes of the system. So that we can obtain a streamlined system with better predictive discrimination performance.

For multi-objective fitness functions, different weight ratios are considered. \( w \) represents the weight of the smallest number of variables which have less influence on the subsequent classification results than the two misclassification rates, so a smaller weight ratio can be set. At the same time, we need to conduct contrast test using the standard binary particle swarm optimization (BPSO, Binary Particle Swarm Optimization) and quantum behaved particle Swarm Optimization (QPSO) [9]-[10], where the population of the two particle swarms and the maximum number of iterations are the same as CQPSO. The optimized iterative experiment was carried out continuously for 60 rounds, and the fitness value and convergence time are taken as the average values after multiple experiments.

Fig. 2 shows the convergence of a certain optimization process for different optimization algorithms. It can be seen that the fitness value of CQPSO is 0.039 and the convergence is best. In terms of convergence speed, QPSO has the fastest convergence and BPSO has the slowest convergence. The iteration speed of CQPSO is moderate. Table 2 lists the mean of the optimal fitness values achieved by the objective function after applying each algorithm, the mean number of iterations for each optimization algorithm, and the number of optimized average variables. The fitness value of CQPSO is smaller than the other two optimization algorithms. The classification optimization model has higher classification accuracy and presents a better global search advantage. Overall, the optimization performance of CQPSO is better than the other two algorithms.
The reference space is reconstructed by using the optimal variable combination optimized by CQPSO algorithm, as shown in Fig. 3. As can be seen from the figure, the optimized variables make the MD value difference between the normal sample and the abnormal sample more obvious, and the overlap is much less than before the optimization.

**Conclusion and Outlook**

In order to solve the problem that the standard binary particle swarm is premature and slow in MTS variable reduction, this paper proposes a multi-objective dimensionality reduction model based on the Mahalanobis distance of Gram-Schmidt orthogonalization and classification threshold, and solves the system with CQPSO algorithm. This method was applied in the diagnosis experiment of steel plate fault defect. On the whole, the overall optimization effect of CQPSO is more ideal.

**Acknowledgements**

The research has been supported by the National Natural Science Funds (61473144).
References

[1] Yang T H, Cheng YT. The use of Mahalanobis–Taguchi System to improve flip-chip bumping height inspection efficiency[J]. Microelectronics Reliability, 2010, 50(3): 407-414 DOI: 10.1016/j.microrel.2009.12.001

[2] Huang CL, Hsu TS, Liu CM. The Mahalanobis–Taguchi system—Neural network algorithm for data-mining in dynamic environments[J]. Expert Systems with Applications, 2009, 36(3):5475-5480 DOI: 10.1016/j.eswa.2008.06.120

[3] Shakya P, Kulkarni M S, Darpe A K. Bearing diagnosis based on Mahalanobis–Taguchi–Gram–Schmidt method[J]. Journal of Sound and Vibration, 2015, 337:342-362. DOI: 10.1016/j.jsv.2014.10.034

[4] Rizal M, Ghani JA, Nuawi MZ. Cutting tool wear classification and detection using multi-sensor signals and Mahalanobis-Taguchi System[J]. Wear, 2017:376(15): 1759-1765 DOI: 10.1016/j.wear.2017.02.017

[5] Yazid Mam, Rijal JK, Awaluddin MS, Sari Emelia. Pattern Recognition on Remanufacturing Automotive Component as Support Decision Making Using Mahalanobis-taguchi System[J]. Procedia CIRP, 2015, 26:258-263 DOI: 10.1016/j.procir.2014.07.025

[6] Zeng Jiang-hui, Zeng Feng-zhang. The Measurement Scale of Mahalanobis-Taguchi System Optimization Based on Fuzzy Robustness Discriminant Criterion [J]. Industrial Engineering and Management, 2008, 13(3):52-55. DOI: 10.19495/j.cnki.1007-5429.2008.03.013

[7] Iquebal A S, Pal A, Ceglarek D, et al. Enhancement of Mahalanobis–Taguchi System via Rough Sets based Feature Selection[J]. Expert Systems with Applications, 2014, 41(17):8003-8015. DOI: 10.1016/j.eswa.2014.06.019

[8] Niu Junlei. Methods of Classification and Sort Evaluation using Mahalanobis-Taguchi System Based on Omni-Optimizer Algorithm and Applications[D]. Nanjing University of Science & Technology, 2012.

[9] Reséndiz E, Moncayo-Martínez, LA, SolísG. Binary ant colony optimization applied to variable screening in the Mahalanobis–Taguchi System[J]. Expert Systems with Applications, 2013, 40(2):634-637. DOI: 10.1016/j.eswa.2012.07.058

[10] Pal A, Maiti J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis-Taguchi system using Mahalanobis distance and binary particle swarm optimization[J]. Expert Systems with Applications. 2010, 37(2):1286-1293 DOI: 10.1016/j.eswa.2009.06.011

[11] Jin X H, Chow T W S. Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis–Taguchi system[J]. Expert Systems with Applications, 2013, 40(15):5787-5795 DOI: 10.1016/j.eswa.2013.04.024

[12] Huang C L, Hsu T S, Liu C M. The Mahalanobis–Taguchi system—Neural network algorithm for data-mining in dynamic environments[J]. Expert Systems with Applications, 2009, 36(3-part-P1):5475-5480 DOI: 10.1016/j.eswa.2008.06.12