

Abnormal Quality Pattern Recognition of Industrial Process Based on Multi-Support Vector Machine

Fengwei Guan¹, Lianglun Cheng^{2*}

¹ College of Computer, Guangdong University of Technology, GuangZhou, 510006, Guangdong, China.

² College of Automation, Guangdong University of Technology, Guangzhou 510006, Guangdong, China.

* Corresponding author. Tel: +86 13602771137; email: guan2556120684@163.com

Manuscript submitted August 2, 2018; accepted September 20, 2018.

doi: 10.17706/jsw.13.9.506-519

Abstract: This paper studies the quality pattern recognition of industrial process based on the statistical process control (SPC). An abnormal quality pattern recognition model based on multi-support vector machine was proposed, which can be used to solve the problem of abnormal pattern recognition in the intelligent manufacturing process for products. The combination of "one-to-one" and "one-to-many" support vector machine (SVM) classifiers is arranged according to the structure of directed acyclic graphs in the model. At the same time, a structural optimization method was proposed to reduce the cumulative error problem. The model uses the original features of the data stream of quality. For the support vector machine classifier with low recognition accuracy, the statistical features and shape features form the data stream of quality are integrated with the original features. Relief algorithm is used to reduce the fusion features in order to reduce the consumption caused by increased features. The experimental results demonstrate that the model improves the accuracy of the recognition of abnormal patterns, and its structure also has a good time advantage.

Key words: Abnormal quality, pattern recognition, SPC, SVM, MSVM.

1. Introduction

Recognizing abnormal quality patterns in industrial manufacturing processes accurately, analyzing and locating anomalies has great significance for industrial production. In industrial manufacturing processes, SPC is usually used to analyze manufacturing process anomalies for improvements and quality of products [1].

The control chart is one of key tools used in SPC. Control chart is a quality management chart based on the principle of mathematical statistics to analyze and judge whether the process is in a stable state. Recognizing the quality patterns with high similarity in quality management charts accurately and fast is a difficult problem. In recent years, many scholars have conducted in-depth research on recognition of quality anomaly pattern based on control charts. Hassan et al. studied the statistical characteristics of the control graph model earlier and proposed a method based on feature selection to recognize abnormal pattern for the process with independent discrete variables [2]. Ranaee extracted the sample data sets and used the subspace of features to recognize abnormal pattern [3]. Cheng H.P. used the correlation characteristics between time series sets to narrow the range of abnormal parameters [4]. Shiyuan Yang proposed an intelligent recognition method for control chart pattern based on principal component analysis (PCA) and

SVM [5]. Gauri summarized seven independent shape features of quality pattern for abnormal quality patterns recognition[6]. Cheng Z used probability neural network to recognize control patterns [7]. Q Le proposes a method that using BP network to recognize the quality patterns [8]. However, the methods based on the feature selection proposed above use the same data stream for all classifiers, so the recognizing effect for each pattern cannot be guaranteed to be best. The method based network has extremely complicated structure due to high dimension of the quality data in manufacturing process.

To our best knowledge, some quality patterns are very similar and difficult to recognize fast in data stream. There are few methods so far which can accurately recognize similar abnormal patterns using quality data streams with good time efficiency[9]. Therefore, this paper proposes a model for abnormal quality recognition based on multi-support vector machine (MSVM), which uses SVM as the basic classifier. The original features are fused with statistical and shape features to improve recognition accuracy for SVM classifier which has low recognition accuracy. Then Relief algorithm is used to reduce the impact of increased dimensions in data stream. Basic SVM classifiers are organized by structure similar to Directed Acyclic Graph (DAG), and classifier structure design principle is proposed at the same time, which can effectively reduce the number of basic classifiers and improve recognition accuracy and time efficiency.

2. Related Works

2.1. The Quality Patterns in Dynamic Data Stream

There are many quality patterns in the industrial manufacturing process. Western Electric has studied the quality anomaly patterns and six basic quality patterns have been defined, which cover the common quality patterns in the intelligent manufacturing process [10]. The rest of the anomaly patterns can be obtained from the combination of these six basic patterns. The six patterns in the actual production process are shown as normal pattern (Nor), increasing trend pattern (IT), decreasing trend pattern (DT), cycle pattern (Cyc), upward shift pattern (US), downward shift pattern (DS).

The data forms of six patterns are shown in Figure 1. The point of the normal pattern on the control chart has a center line that coincides with their mean value. The process of the normal pattern is stable and controlled and only affected by random error and the data distributes as normal distribution which limited with control boundary. Each abnormal pattern shows a specific abnormal shape that marking the process is abnormally affected by some non-random factors and out of control. Among them, DT, DS, IT, and US are very similar in expression, and it is difficult to recognize them in data stream.

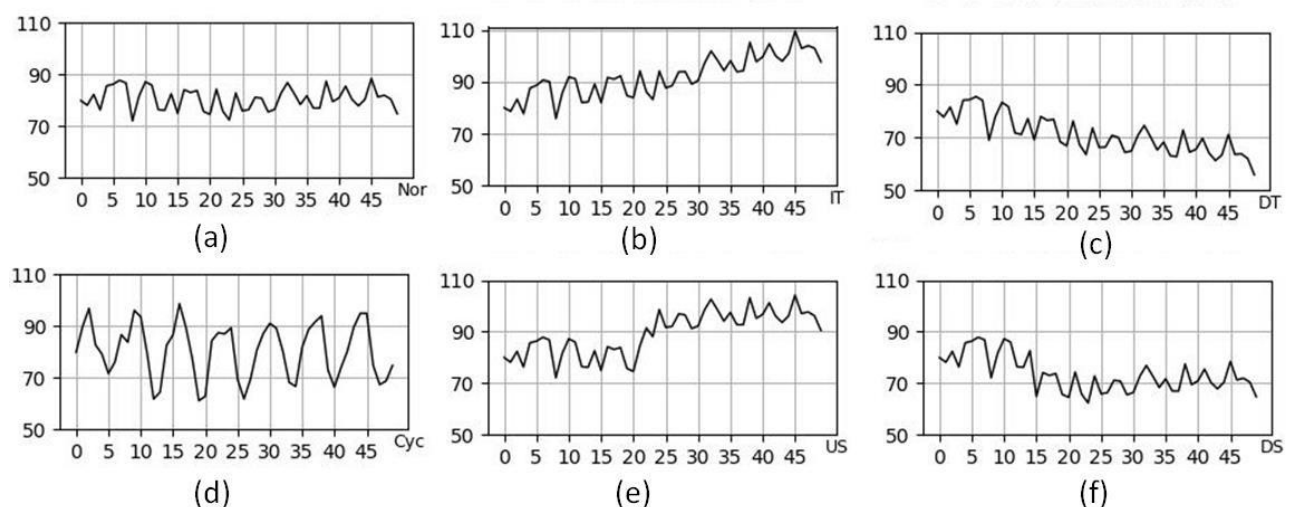


Fig. 1. Control chart of quality patterns in dynamic process: (a) Normal pattern (b) Increasing trend pattern (c) Decreasing trend pattern (d) Cycle pattern (e) Upward shift pattern (f) Downward shift pattern

2.2. Feature Fusion and Reduction Based on SPC

In order to improve the recognition accuracy of the six quality patterns, it is necessary to select the features in control chart to improve classification performance. Features are selected from the original features as classification features directly or extracting statistical and shape features from the original data as classification features. Different types of classification features on the SPC control chart depict each pattern from different angles. The morphological characteristics of patterns can be highlighted by mapping or transforming the original data on the control chart, which makes the pattern classification easier and the recognition accuracy is better.

Common feature extractions include statistical features and shape features. The statistical features are composed of multiple statistics of the original data stream on the control chart, which can characterize the essence of the patterns and strengthen the difference between different patterns. The shape features are a digital quantity calculated on the control chart, which can effectively characterize the geometry of patterns, and highlight the discrimination between different patterns. In addition to the original features, the most commonly used 8-dimensional statistical features and 4-dimensional shape features are selected, which can be merged into the original features to enhance the recognition accuracy of each pattern. Those names and calculation equations are shown in Table 1:

Table 1. The Statistical and Shape Characteristics

	Feature name	Calculation equations		Feature name	Calculation formula
1	Mean	$X_{mean} = \sum_{i=1}^N x_i / N$	7	Pulse (I)	$X_I = X_{max} / X_{amp}$
2	Standard deviation (SD)	$X_{sd} = \sum_{i=1}^N (x_i - x_{mean})^2 / (N-1)$	8	Waveform (K)	$X_k = X_{ms} / X_{amp}$
3	Skewness (SK)	$X_{skewness} = \left[\frac{\sum_{i=1}^N f_i (x_i - x_{mean})^3}{N} \right] / s^3$	9	SB	$SB = \text{sgn} \left[\frac{\sum_{i=1}^N y_i (t_i - \bar{t})}{\sum_{i=1}^N (t_i - \bar{t})^2} \right]$ ($t_i = ic (i = 1, 2, \dots, N)$)
4	Kurtosis (Kur)	$X_{kurtness} = \sum_{i=1}^N (x_i - x_{mean})^4 / N (X_{sd})^4$	10	PSMLSC (PS)	$PSMLSC = \sum_{i=1}^{N-1} (O_i - O_i') / 2N$
5	Mean square amplitude (MS)	$X_{ms} = X_{max} / X_{kurt}$	11	SRSNGE (SR)	$SRANGE = \max(s_{jk}) - \min(s_{jk})$
6	Dispersion (Dev)	$X_{devsq} = \sum_{i=1}^N (x_i - \bar{x})^2$	12	REAE	$REAE = \left[\frac{MSE}{MSE_{jk} / 6} \right]$ $j = 1, 2, 3; k = 2, 3, 4; j < k$

Feature fusion increases the dimension of data stream. The increase of the feature dimension makes the estimation of statistical property difficultly, which leads to the decline of the generalization ability of the classifier. Therefore, it is necessary to adopt a certain method to reduce the feature dimension of data stream.

The Relief algorithm is a features selection algorithm based on features weights. The classification decision surface is used by Relief which is similar to the support plane of SVM. So Relief is considered as an improvement of SVM classifier's recognition ability. Therefore, this paper use Relief algorithm to select the importance features. Relief uses the classification idea of "hypothesis interval", maximizing the maximum distance that the decision surface can move while keeping the classification of samples unchanged. The

distance θ can be expressed as Equation 1:

$$\theta = \frac{1}{2} (x - M(x) - x - H(x)) \quad (1)$$

In Equation 1, $M(x)$ and $H(x)$ refer to the nearest neighbors of same class as x and the different class to x . When a feature is advantageous for classification, the distance of the same samples on the feature is relatively close, while the distance of the different classes samples on the feature is relatively long, corresponding to the smaller of the first term and the larger of the second term in the parentheses of the Equation 1. The "hypothesis interval" can evaluate the classification ability of all features, so that the features subsets most useful for classification can be approximated. The relevant statistic corresponding to the features j is as shown in Equation 2:

$$\delta^j = \sum_i -diff(x_i^j, x_{i,nh}^j)^2 + diff(x_i^j, x_{i,nm}^j)^2 \quad (2)$$

Where x_i^j represents the value of the sample x_i on feature j ; $x_{i,nh}$ and $x_{i,nm}$ represent the nearest point of the same class and different classes with x_i respectively. Therefore, the larger the correlation statistic of feature j , the more important the feature is. For the introduced 12-dimensional auxiliary features, the top 8 most important features are selected to reduce features.

3. Algorithms

3.1. SVM Classifier

SVM computational complexity is not affected by dimension of the features. In the case of high dimension data stream, it also has a good classification ability and computational feasibility. Therefore, SVM classifier is selected as the basic classifier to reduce the impact that caused by increasing and high dimension of data stream.

SVM usually includes linear separable SVM and nonlinear SVM. In order to make SVM more general, the slack variable ξ_i and penalty factor C are introduced in linear separable constraints. So the optimization for the hyper-plane of classification under completely linear separate condition converts to Equation 3:

$$\min \frac{1}{2} w^2 + C \sum_{i=1}^N \xi_i \quad \text{s.t. } y_i (wx + b) \geq 1 - \xi_i, \quad i = 1, \dots, n \quad (3)$$

The solution of the optimization for the classification hyper-plane is a convex optimization problem, so the Lagrange function and the kernel function $k(x_i, x)$ can be introduced to obtain the target equation shown in Equation(4):

$$f(x) = \text{sgn} \left| \sum \alpha_i^* y_i k(x_i, x) + b^* \right| \quad (4)$$

Kernel function has great significance on the recognition effect. The kernel function defines how to map low-dimensional data and avoids the dimensional explosion of linear mapping. Common kernel functions include Nystroem kernel function, Fourier kernel function, linear kernel function and Gaussian RBF. The RBF kernel function has excellent generalization ability in high-dimensional data. Its definition is as follows:

$$\text{RBF}(x_i, x_j) = \exp \left(-\frac{|x_i - x_j|^2}{\sigma^2} \right) \quad (5)$$

There are two hyper-parameters including σ and penalty parameter C when RBF is selected as a kernel function. The parameter σ describes the data distribution or range characteristic of the training samples and determines the width of the local domain. Another hyper-parameter C represents a penalty for misclassification, achieving a compromise between the proportion of misclassified samples and the complexity of the algorithm. Therefore, SVM applies to two classification problems especially when the data dimension is high.

3.2. Multi-class SVM

Recognition of dynamic process abnormal quality patterns is a multi-classification application. MSVM is usually used to solve this problem. The more commonly used MSVM includes one versus rest SVM (OVR SVM), one versus one SVM (OVO SVM) and directed acyclic graph SVM (DAG SVM).

The OVR SVM is the earliest proposed multi-class SVM. OVR SVM needs K SVM classifiers for the K classification problem. Each SVM classifier is responsible for recognizing one of the patterns from the others, so the OVR SVM only requires K SVM classifiers totally. Therefore, OVR SVM has advantage in the number of classifiers and high classification efficiency. However, each basic classifier in OVR SVM needs all training sets as training samples, and the solution complexity of OVR SVM increases exponentially with the increasing of samples number. OVR SVM has situation of "rejection" and "re-division", and there is a strong classification tendency in the case of samples asymmetry.

The OVO SVM will combine all classes and then turn them into two classification problems. It needs $k(k-1)/2$ SVM classifiers to recognize the K patterns. Each sample is entered into all basic SVM classifiers for recognizing, and the pattern with the highest number of votes in the final result is used as OVO SVM's output. OVO SVM can avoid the "rejection" problems, and the training size is much smaller than the OVR SVM, so the OVO SVM training time is also very short. However, OVO SVM is seriously affected by a basic SVM classifier, and it is prone to be over-fitting. When the number of votes is consistent, the classification result cannot be given, and the classification efficiency decreases as the patterns increases.

DAG SVM is an improvement of OVO SVM. Recognizing with K classification problems needs number of $k(k-1)/2$ classifiers. Those classifiers are combined into a directed acyclic graph structure instead of the linear arrangement structure of OVO SVM. The $k(k-1)/2$ two-class SVMs are arranged into triangle with $K-1$ layer, and the bottom K leaf nodes correspond to K quality patterns. The DAG SVM classify samples from top to bottom according to the current two-class SVM classification result until the classification result is obtained. DAG SVM is "connected" from the top node to any leaf node, and always gets and can only get one classification result avoiding the problem of "rejection" and "class overlap". The number of $K-1$ classifiers is needed to classify K patterns. The efficiency of DAG SVM is greatly improved compared with OVO SVM, especially in the case of a large number of patterns.

However, the number of classifiers in DAG SVM is still large, and the arrangement structure of each two-class SVM has a great influence on its classification performance. There is no method to define the structure of DAG SVM currently, and the actual DAG SVM is often a randomly determined structure. The same sample may produce different classification results because different structures of DAG SVM have different classification paths.

3.3. Structure Design of MSVM

Aiming at the problems existing in the current MSVM classifier, this paper proposes a new MSVM recognition model that combines OVR SVM and OVO SVM classifiers. Those classifiers are arranged in a structure similar to the structure of DAG SVM. The structure of MSVM is shown in Figure 2. An OVR classifier with the best performance is used in the first step of recognizing. The remaining $K-1$ patterns continue to be classified by the DAG SVM structure. The number of OVO classifiers is effectively reduced in the MSVM

model.

In addition, the traditional Multi-class SVM model based on features fusion use the same input data stream. The data stream contains the same weight characteristics for each basic SVM classifiers. Although this method is relatively simple to calculate, it cannot guarantee that each OVO classifier achieves the best classification effect. It is unreasonable to use the same feature fusion on the input data stream of all classifiers. The MSVM model proposed in this paper uses different fusion features and parameters for the SVM classifier with low classification accuracy. For those OVO classifiers whose original feature classification effect is higher than the threshold, the original features are only used to train and classify in order to reducing the time.

Since the SVM classifier arrangement has a great influence on the classification results, a method for adjusting the structure of the MSVM is proposed based on the classification effect of OVO. In the MSVM model, the classifier which appears closer to the upper layer is more sensitive to classifier errors. For example, the structure in Figure 2 is one of the structures of MSVM, the first layer is an OVR SVM classifier of 1-R, and the remaining layers are OVOSVM classifiers. Pattern 6 first appears in the second layer of the MSVM model in the structure, so there is only one path for pattern 6 to be correctly classified, and any classifier error in the path will cause the classification to fail. Pattern3 appears for the first time on the third layer, and there are 4 correct classification paths. Therefore, pattern3 is less sensitive to classifier errors. According to this method, we adjust the SVM classifier with poor classification performance to the lower level as much as possible, and reduce the probability of misclassification and the cumulative error.

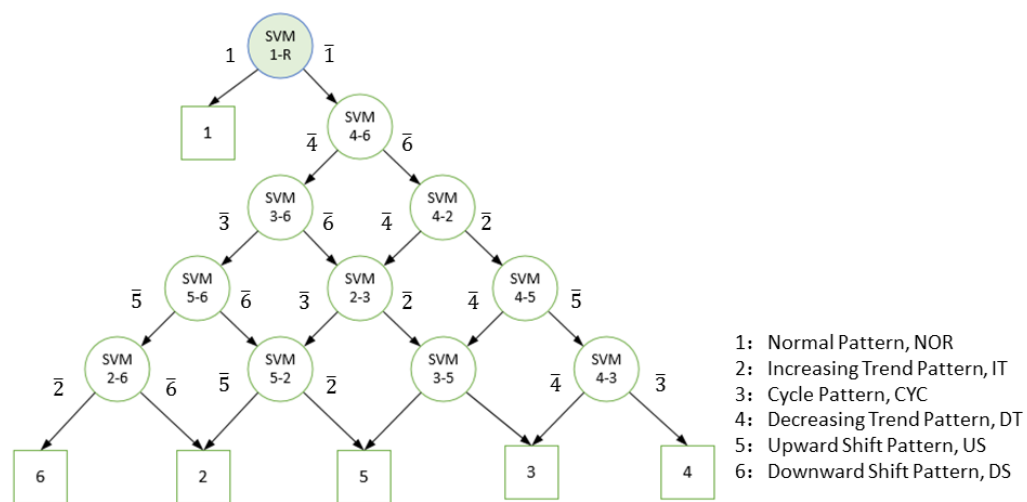


Fig. 2. The structure of MSVM

In general, the MSVM model this paper proposed can reduce K-2 OVO classifiers comparing with DAGSVM. When there are many classification patterns, it can effectively reduce the time and number of training SVM. We also propose a structural design principle of MSVM model which reduces the recognition error. In the structure of MSVM, classifiers use different parameters and only OVO classifiers with low classification accuracy need feature fusion. This can improve time efficiency and recognition accuracy.

4. Experimental Results

In order to verify the feasibility and accuracy of the proposed algorithm model, experiments with the proposed algorithm has been conducted and the result has been compared with other algorithms in this chapter. At the same time, the experimental results were analyzed and discussed. The experimental environment of this paper is windows_10 system with the 8G memory and Intel CPU of I5-8250, it is

programmed in Python, and the compilation environment version is python 3.6.

4.1. Simulation of the Multivariate Process

Since the abnormal pattern data is difficult to obtain in the real process, the research of process patterns recognition is usually based on the data stream simulated by the Monte-carol (MC) method. MC method is a numerical calculation method guided by probability and statistics theory. It uses a pseudo-random number to generate a quality data stream that is similar to the abnormal quality in the actual production process. Therefore, this paper uses data stream generated by MC method to estimate the performance of MSVM.

The sensors in industrial production processes collect data every interval, the interval is usually set as 1 minute. So the data stream with a time length of 60s has been generated by MC. The curve of quality data stream will fluctuate randomly around the design target value under controlled conditions. The quality expression $x(t)$ obeys the normal distribution in the normal pattern and the outliers are randomly introduced during the sampling period according to the type of the exception in the other patterns. In summary, the data stream of the six patterns is generated according to the expressions defined in Table 2. The value of μ and σ in generation equations is set to 80 and 5.

Table 2. Generation of Data Stream

Number	Name of Patterns	Generation Equations	Parameter Values
1	Normal Pattern, NOR	$Y(t) = \mu + \gamma \times \sigma$	$\mu = 80, \sigma = 5$
2	Increasing Trend Pattern, IT	$Y(t) = \mu + \gamma \times \sigma + k \times t$	
3	Decreasing Trend Pattern, DT	$Y(t) = \mu + \gamma \times \sigma - k \times t$	$k \in [0.2, 0.5]$
4	Cycle Pattern, CYC	$Y(t) = \mu + \gamma \times \sigma + A \sin(2\pi t/T)$	$A = 15, T \in [4, 14]$
5	Upward Shift Pattern, US	$Y(t) = \mu + \gamma \times \sigma + g \times s$	$\begin{cases} g = 1 & \text{when } t \in [15, 25] \\ g = 0 & \text{else} \end{cases}$
6	Downward Shift Pattern, DS	$Y(t) = \mu + \gamma \times \sigma - g \times s$	$s \in [7.5, 20]$

In the above way, we use the number of each mode as our recognition target. More than 95% of the data is normal production quality data in the data set of real industrial production, and the other five patterns only account for less than 5%, so the real data set is unbalanced. In order to avoid our model being affected by the problem of imbalance, we generated 10,000 data records for each pattern. Because the data stream in the intelligent manufacturing process is usually high-dimensional, each of cases has 250 features, so we have 60,000 cases of dataset which case has 250 features.

4.2. Kernel Function and Hyper-parameter of SVM

After getting the data set, we need to set and optimize the parameters of the basic SVM classifier to ensure the best recognition accuracy. The kernel function of SVM has a great influence on recognition accuracy. This paper compares common kernel functions in the process of quality pattern recognition.

The kernel functions including Nystroem kernel function, Fourier kernel function, linear kernel function and Gaussian RBF function are compared in time and accuracy. The results are shown in Fig.3, Fig. 3 (a) shows the comparison of training time of different kernel functions while Figure 3 (b) shows the comparison of classification accuracy of different kernel function.

From the time perspective, the training time of linear kernel function and Gaussian RBF function is constant and the Gaussian RBF has less training time. The training time of Nystroem kernel function and Fourier kernel function increases as the number of features increases. When the number of features is greater than 245, the RBF kernel function has the advantage of training time on recognition. In the actual industrial production process, the dimension of the data stream is often higher than 240. So the RBF kernel function usually performs well in training time.

From the perspective of classification accuracy, the accuracy of linear kernel function and Gaussian RBF function is constant with the increasing number of features. But the accuracy of Nystroem kernel function

and Fourier kernel function is gradually improved with the features increasing, and will not increase after reaching a certain value. The solid line parallel to the vertical axis in Figure 3(b) indicates the number of our features. The RBF calculation time is better than other kernel functions at the solid line. Therefore, it is reasonable for the MSVM to select the RBF kernel function as the mapping function.

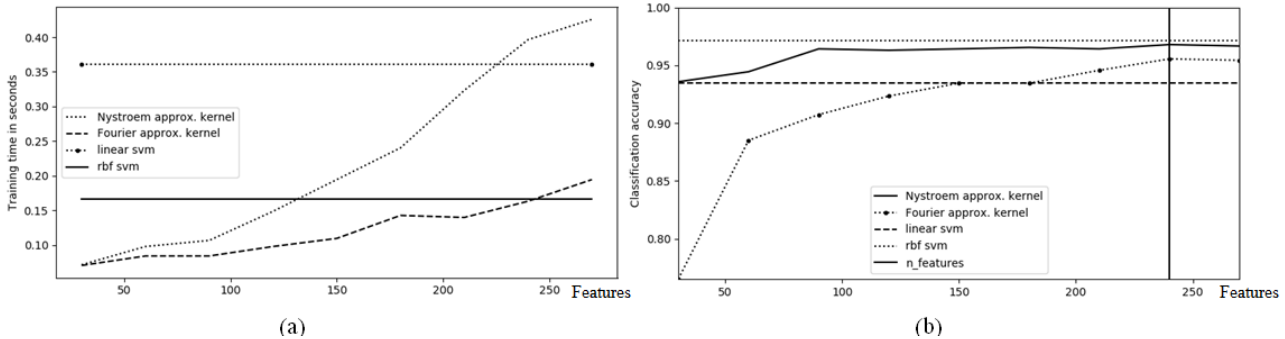


Fig. 3. Selection of kernel function (a) Comparison of training time of different kernel functions (b) Comparison of classification accuracy of different kernel functions

The SVM of RBF kernel function has two important parameters to optimize. The two parameters are σ and C , we have initially determined their range that the model has better classification accuracy. The value of σ is from 1×10^{-5} to 1×10^{-1} and the range of C is from 0.1 to 6.4. The grid search algorithm can find the optimal combination of parameters in the specified space. It is a commonly used algorithm for parameter optimization. So we use grid search algorithm to find the optimal combination of parameters. The optimization results are shown in Figure 4. In the process of parameter optimization, we found that SVM can achieve good recognition accuracy when σ is 0.015 and the value of C is 0.7. The experiment in this chapter using the parameters we have optimized. All of the experimental results have been verified by a 3-fold cross-validation.

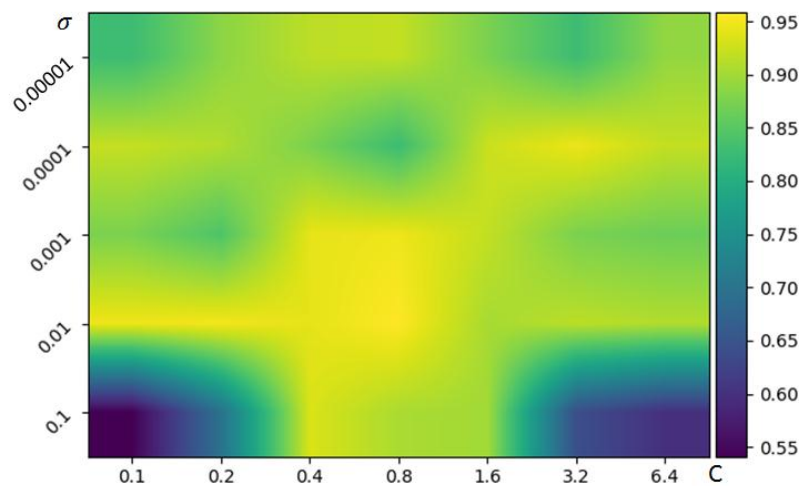


Fig. 4. Optimization of SVM parameters

4.3. Features Selection and Fusion

After the SVM parameters are determined, we need to find out which statistical features and shape features are used to promote the recognition of different patterns, and then use these selected features to improve our recognition accuracy. There are six different OVR recognition requirements, which are named

as 1-R, 2-R, ... 6-R by the pattern number in Table 2. For example, the '1-R' means to recognize the NOR pattern and the other five patterns.

Table 3 the Important Features for OVR SVM

Classifier	Pattern	Important features
1-R	Nor-R	Mean, Kur, Dev, K, RE, SD, MS, SR
2-R	IT-R	SR, Mean, MS, RE, PS, SD, Kur, Dev, I
3-R	CYC-R	PS, MS, I, Kur, SK, Dev, PE, SB
4-R	DT-R	SB, SK, Kur, PS, MS, I, RE
5-R	US-R	Kur, Mean, SR, I, SK, RESE, K, PS
6-R	DS-R	Kur, Mean, SD, SR, Dev, Kur, SK, I, PS

For each of the six patterns, we adjust the label of the dataset, marking data of each pattern as label '0', and the other five patterns are labeled as '1'. Then we use the Relief algorithm to select the most important first 8 features from the 12 statistical features and shape features appearing in Table 1. The results of the features selection are shown in Table 3. Abnormal patterns will cause the mean of data to change except CYC pattern. So the mean value promotes the correct classification of 1-R, 2-R, 5-R and 6-R patterns well. The kurtosis (Kur) characterizes the features of the probability density distribution curve at the peak value, and it can be used to distinguish US-R, DS-R well. SR and SB play an important role in recognizing IT-R and DT-R patterns.

The Relief algorithm can also calculate important coefficients while giving important features. We also analyzed and compared the characteristic importance coefficient of the OVR recognition. The distribution of 4-R is representative, so we plot the coefficient proportional histogram of 4-R in Fig 5. PS has the most important contribution to recognize CYC pattern from the others. MS, I and K also are very important features for recognition.

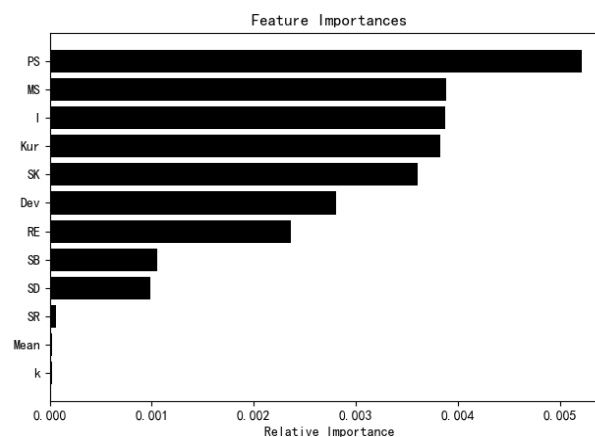


Fig. 5. Importance features of 3-R patterns

After completing the feature selection for the OVR classifier, we select features for the remaining 10 OVOSVMs. In the process of feature selection of OVO classifiers, some meaningful conclusions have been found in Experiments. Kur responses the overall curvature of the quality curve and can be used to recognize NOR pattern and shift patterns effectively. Mean square amplitude (MS) performances better between CYC pattern and trend pattern. Dispersion (Dev) has a higher contribution in recognizing NOR pattern and CYC pattern. Waveform (K) uses the difference of frequency and amplitude to describe the curve of different displacements of each data point at the same time. Therefore, the recognition between trend pattern and shift patterns is much better. SB can represent the slope of the least square line of the quality curve, and the

slope of the increment trend(IT) and the upward shift pattern is greater than zero, but the downward trend(DT) and the downward shift(DS) pattern is less than zero. The slope of the CYC mode and NOR mode curve is relatively smooth. Therefore, the feature can recognize the increment trend(IT) and the downward trend (DT) from the upward shift(US) and the downward shift(DS). PSMLSC indicates mean ratio of points where the mean line and the least square line are intersected, and PSMLSC has a higher discrimination between normal (Nor) pattern and trend pattern. The performance of SRSNGE recognition on step pattern is higher than trend pattern while the NOR pattern is also lower than CYC pattern.

There are 8 features selected from 12 statistical and shape features for the OVR and OVO classifiers. The features are arranged in non-incremental order according to importance. When training the SVM classifier, the fused features are sequentially connected to the original features as training data stream.

4.4. Accuracy Analysis of MSVM Model

After completing the optimization of parameters and the selection of features, we will set the parameters of the SVM and incorporate the selected features into the original data to verify whether the addition of these features can improve the classification accuracy of the model.

In order to reduce the impact of accidental factors in the experiment, we split the data stream into 5 batches for the OVR and use them for cross-validation. Each batch of data contains a set of 2000 positive cases and 10000 negative cases. The experimental results were averaged of 10 times. The original feature recognition accuracy of the six patterns and the recognition accuracy after the features fusion are shown in Table 3. Classifier of DS-R has the highest recognition accuracy of 95.55% using raw data for pattern recognition. The recognition accuracy of DS-R is relatively low, only 93.33% before using the fusion features. When the statistical features and shape features with higher classification contribution are combined with original features, the classification ability of the OVR SVM classifier improve obviously, especially the classification performance of 2-R classifier reaches 98.7%.

Table 3. Comparison of Precision for OVR SVM

Pattern	Accuracy only Using Original Features	Accuracy after Adding Selected Features into Original Features
Nor-R	93.96%±0.3%	95.88%±0.51%
CYC-R	95.2%±0.26%	99.7%±0.46%
IT-R	95%±0.4%	96.4%±0.6%
DT-R	93.33%±0.25%	94.44%±0.86%
US-R	95.0%±0.62%	95.5%±0.48%
DS-R	95.55%±0.27%	96.66%±0.82%

In this paper, the original data stream is also used as input for 10 OVOSVM classifiers. The Decision Tree (DT), Random Forest (RF), XGboost (GB) and SVM using only raw features classifiers are basic algorithms. The comparison of precision between basic algorithms and MSVM are shown in Table 4. Logistic Regression has a best accuracy of 99.05% when recognize the DS-IT pattern. Random forest results are not as good as SVM that using raw data, but better than other algorithms. We better not use decision trees to recognize DS-DT pattern because the accuracy is only 71.52% and it is the lowest accuracy in the basic algorithms. Gradient Boosting has a moderate accuracy in those algorithms. Experimental results show that the recognition accuracy of SVM using only raw features is higher than other classifiers, which verifies the feasibility and rationality of selecting SVM as the basic classifier. The minimum accuracy of SVM raw features is 90.33% for US-IT pattern, and the highest accuracy is 99.05% for DS-IT pattern.

Among OVO classifiers of DS-NOR, DS-IT, DS-US, US-NOR, US-DT, DT-NOR and DT-IT, the classification accuracy is all over 97%. To reduce the time consumption, only the original features are used to recognize these patterns instead of adding fused features. We believe that the original feature method can ensure the

accuracy of the pattern recognition and minimize the time consumption. This makes our model can be applied to recognize online quality data streams in real time. The DS-DT, US-IT, and IT-NOR classifier recognition accuracy have room for improvement. Therefore, we performed feature fusion experiments on these three classifiers. After merging the selection features, we found that the accuracy of SVM classifiers has been improved. The accuracy for DS-DT pattern has been improved from 86.33% to 90.82%. US-IT pattern has been improved from 93.33% to 95.06% and the accuracy of IT-NOR pattern improved 2.12%. The recognition accuracy improves from 2.04% to 3.49% after using the fusion of features for the DS-DT, US-IT and IT-NOR patterns. Therefore, the fusion feature can improve the recognition accuracy of the classifier.

Table 4. Accuracy comparison of each classifier

Accuracy (%)	Logistic Regression	Random Forest	Decision Tree	Gradient Boosting	Raw features SVM	Fusion Features SVM
DS-NOR	89.23	96.15	88.56	87.31	98.50	/
DS-IT	99.05	99.00	97.54	97.96	99.05	/
DS-DT	77.45	84.68	71.52	79.00	86.33	91.82
DS-US	98.61	99.00	94.96	94.50	98.50	/
US-NOR	90.27	98.40	86.26	96.37	97.35	/
US-IT	83.33	86.70	80.02	90.41	93.33	95.06
US-DT	91.01	97.95	94.05	98.05	97.52	/
DT-NOR	94.20	97.66	88.04	90.03	98.31	/
DT-IT	95.65	98.74	95.30	97.08	98.00	/
IT-NOR	94.30	96.43	96.67	96.67	96.24	98.33

In this paper, the feature fusion of DS-DT, US-IT, and IT-NOR classifiers is performed. The experimental accuracy is recorded in the MSVM column in Table 4. The process of accuracy change in the feature fusion is shown in Figure 6. With the fused features been adding to the original features one by one, the recognition accuracy of the SVM classifier is gradually improved. After the 4 or 5 features are combined, the recognition accuracy of the classifier reaches the peak. The accuracy of IT-NOR pattern is up to 98.33%. US-IT pattern gets the highest classification accuracy of 95.06% after adding the sixth fusion feature. DS-DT pattern has a best accuracy of 91.82% after adding five fusion features. With the increase of the fusion features, the recognition accuracy is no longer improved, but a small decrease occurs, and the model appears to be over-fitting.

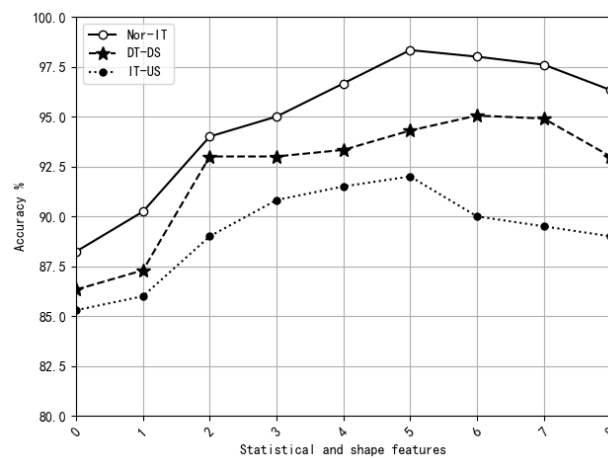


Fig. 6. OVO classification accuracy with features fusion

4.5. Structure Optimization for MSVM

According to the accuracy of each SVM classifier, we adjust the structure of the MSVM to reduce the cumulative error and improve the classification efficiency of the model. The adjusted classifier arrangement structure is shown in Figure 7. The experiment demonstrates that the CYC pattern (4-R) has the best recognition accuracy after combining original with statistical features and shape features. According to the method of MSVM structure adjustment proposed in this paper, the 4-R classifier is selected as the first layer of MSVM. The CYC pattern is recognized from the others firstly. The recognition accuracy of DS-DT, US-IT is lower in them. According to the MSVM structure adjustment method, the 6-3(DS-DT) classifier and 5-2(US-IT) classifier are adjusted to last level. We have already marked them in Figure 7. The MSVM model proposed in this paper can minimize the cumulative error and ensure the accuracy of recognition through the structural adjustment.

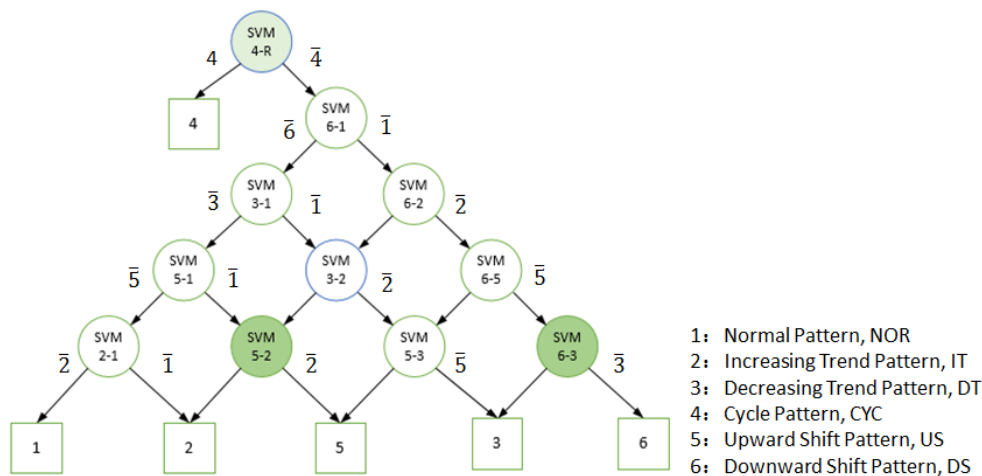


Fig. 7. The structure of MSVM after optimization

4.6. Analysis of MSVM Time Complexity

We also study the time consumption of commonly used SVM algorithm in the field of industrial process pattern recognition. The experimental results are shown in Figure 8. The horizontal axis is the number of patterns, and the vertical axis is the time cost of the algorithm. We found that the time consumption of the OVOSVM and DAGSVM algorithms is consistent, and their time consumption curves are coincident in Figure 8. When the number of recognition patterns is less than 13, OVR SVM consumes more time than other comparison algorithms. When the number of recognition patterns is less than 15, the MSVM algorithm proposed by this paper is the smallest in terms of time consumption, so the MSVM algorithm has a time advantage.

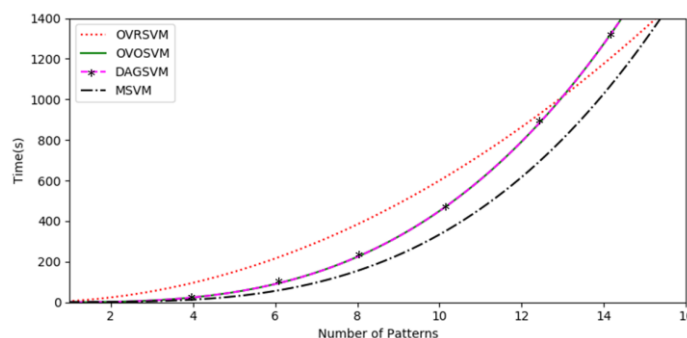


Fig. 8. Time consumption of Multi-class SVM classifiers

The MSM model proposed in this paper only needs 1 time for OVR training and 10 OVO trainings in total. Four training times of OVO classifiers have been reduced compared with the DAGSVM, which greatly shortens the training time of the model. The training time is reduced by K-2. When the recognition pattern is increased, the time advantage of the model will be more obvious. In addition, there are only 3 OVO classifiers needs feature fusion processing in total 10 OVOs for recognizing 6 patterns, which reduces the fusion characteristics of 7 OVO classifiers compared with the traditional 10 OVO classifiers. In summary, the model presented in this paper has a more time advantage.

5. Conclusion

In this paper, an MSVM quality anomaly recognition model is proposed, which focuses on the quality anomaly patterns in the intelligent manufacturing process. The main contribution in this paper is that we propose a MSVM model based on feature fusion for pattern recognition in industrial process quality anomalies. This recognition model consists of OVR SVM and OVOSVM classifiers. The model uses the fusion of statistical features and shape features for SVM classifiers with low recognition accuracy. The Relief algorithm has been used to fuse features for reducing the dimension of features. At the same time, this paper proposes a structural optimization method to reduce the cumulative error in MSVM. In the process of building an MSVM model using this method, the OVR classifier which has the best classification performance is adopted in the first step of classification. The subsequent steps are combined by OVOSVM classifier in the form of DAGSVM classification. The method is making the classifier with low accuracy be in the layer close to the leaf node to avoid the cumulative error of the MSVM model. The model of MSVM can be used in the intelligent manufacturing process to recognize abnormal quality patterns. The experimental results demonstrate that the proposed model has good accuracy and time advantage.

The future work of this paper is to continue to optimize the model structure of MSVM, and try to use the recently proposed algorithm such as deep learning in the recognition of the industrial process quality anomaly pattern. At the same time, there are some limitations in the research of this paper. The amount of data in our simulation is not comprehensive enough to contain all the abnormalities of the actual industrial process. The research in this paper still needs to be further expanded and deepened, and applied in the actual production data flow.

6. Acknowledgment

This work was sponsored by Research and industrialization of intelligent manufacturing integration platform for marine products, Guangdong Provincial Science and Technology Plan Project (grant number 2016B090918045).

References

- [1] Qian, Y. X., & J. Jianfen. (2004). Research of statistical process control (SPC) and its application. *Computer Engineering*, 30(19), 144-145.
- [2] Hassan, A., NabiBaksh, M. S., Shaharoun, A. M. *et al.* (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research*, 41, 1587-1603
- [3] Control chart pattern recognition using an optimize Ebrahimzadeh A, Ranaee V. Control chart pattern recognition using an optimized neural network and efficient features. *Isa Transaction*.
- [4] Cheng, H. P., & Cheng, C. S. (2011). Artificial immune algorithm-based approach to recognizing unnatural patterns among auto correlated characteristics. *African Journal of Business Management*, 5(16), 6801-6813.
- [5] Yang, S. Y., De-Hui, W. U., & Hai-Tao, S. U. (2006). Abnormal pattern recognition method for control

chart based on principal component analysis and support vector machine. *Journal of System Simulation*, 18(5), 1314-1318.

- [6] Gauri, S. K., Chakraborty S. feature-based recognition of control chart patterns. *Pergamon Press*, Inc.
- [7] Cheng, Z., & Ma, Y. Z. (2008). A research about pattern recognition of control chart using probability neural network. *Proceedings of the Isecs International Colloquium on Computing, Communication, Control, and Management*.
- [8] Le, Q., Gao, X., Teng, L. *et al.* (2004). A new ANN model and its application in pattern recognition of control charts. *Proceedings of the Intelligent Control and Automation*.
- [9] Hassan, A., Shariffnabibaksh, M., Shahraroun, A. M., *et al.* (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research*, 41(7), 1587-1603
- [10] Guh, R. (2013). Robustness of the neural network based control chart pattern recognition system to non-normality. *International Journal of Quality and Reliability Management*, 19(1), 97-112.
- [11] Gauri, S. K. (2006). Chakraborty S. Feature-based recognition of control chart patterns. *Pergamon Press*, Inc. 2006.
- [12] Benov, D. M. (2016). The manhattan project, the first electronic computer and the monte carlo method. *Monte Carlo Methods and Applications*, 22(1), 73-79



Fengwei Guan was born in HeBei province, China, in 1992. He graduated in automation from LiaoNing University of Technology in 2016. He is a master's student in school of computer of Guangdong University of Technology. His current research is about pattern recognition of industrial process, data mining.



Lianglun Cheng was born in Hubei province, China, in 1964, is a Prof. and doctoral supervisor of Faculty of Automation, Guangdong University of Technology. He received PhD degree in Huazhong University of Science and Technology. His current research is about computer vision, image data compression etc.