Algorithm of multi-causes process quality diagnosis based on binary tree and intermediate quality indices

Niu qing

Department of industrial design, lanzhou jiaotong university, lanzhou, gansu, p.r.china

Emai: niuqing2000@126.com

Key words: processes quality diagnosis, binary tree, intermediate quality index, cause-selecting control chart

Abstract. To solve the problem of multi-causes process quality diagnosis in manufacturing, an algorithm based on the binary tree of quality cause set and intermediate quality indices is proposed. Firstly, the set of quality cause is divided progressively to form a complete binary tree, and then every node of the subsets is chose as non-control cause set to construct the cause-selecting control chart of the intermediate quality index in its parent node. In quality diagnosis, control charts with same parent are compared to narrow the scope of the abnormal quality causes until they are confirmed eventually. Practice proves that the new algorithm proposed in this paper could make the procedure of quality cause judgment independent to the adjustment one, and the comparison times of control charts could be greatly reduced, so the efficiency of diagnosis is highly promoted.

Introduction

with the development of modern global market, product's quality has been one of the important factors which greatly influence the enterprise competitiveness. Process quality is the foundation of product's quality because the product's quality would be influenced by every process quality directly or indirectly, so process quality control is the core of quality management in manufacturing.

much works had been done on process quality control. Pham and waini [1] proposed an on-line detection model of control charts based on feature extraction and analysis, yu and wu [2] proposed a bayesian-based algorithm of quality control in small batch manufacturing, and a standardized control chart of process mean and variance based on probability integral transformation was suggested by miao et al [3]. These methods could give a judgment that whether the process quality is normal, however, they can not diagnose the abnormal quality causes efficiently when the process quality is out of control.

in the field of process quality diagnosis, method based on two kinds of quality control charts was firstly invented by zhang [4], and improved by zheng [5] in order to reduce the comparison times of control charts on probability. In above theories, once an abnormal cause is confirmed, it must be adjusted to the normal level should the next step of diagnosis be executed, so the period of diagnosis is long. After that, an algorithm based on fuzzy relation model was put forward by liu and jiang [6]. Hyun, kwang and myong [7] proposed an empirical model-based diagnosis framework for batch manufacturing. In another way, with the development of artificial intelligence, neural networks had been introduced to this field and a lot of network models aimed at special applications were founded [8-14]. However, these theories lack of strict mathematical model because the mapping relationship between process quality and its causes set is constructed by experience.

based on the above theories, this paper proposed an algorithm based on constructing the binary

tree of quality cause set and the introduction of intermediate quality indices. Practice proved that the new algorithm could reduce the comparison times of control charts greatly. Moreover, abnormal cause need not be adjusted on the diagnosis procedure so that the time needed is shortened and the efficiency is highly increased.

Basic methods of process quality diagnosis

suppose y is the process quality index and influenced by the quality causes $x_1, x_2, …, x_n$, the total control chart of y has been established in stable state to monitor the process quality. According to the diagnosis theory of two kinds of control charts, assume x_i is selected as a non-control cause and the corresponding cause-selecting control chart has been done, so when the total control chart of y is out of control and meanwhile the cause-selecting control chart about x_i is under control, conclusion can be drawn that only x_i is the abnormal cause. When the two charts are all anomalous, it reveals that other causes except x_i are abnormal, and whether x_i is normal need to consult its control chart.

Diagnosis method based on two kinds of control charts

the basic idea of diagnosis method based on two kinds of control charts is adding one quality cause to non-control causes set by steps, and establishes the cause-selecting control chart of y about the set. In diagnosis, according to the addition sequence of non-control causes, determine whether the current cause is normal. When the current cause is confirmed abnormal, it must be adjusted to the normal level, and then the next one could be diagnosed until all the abnormal causes are found. The basic principles of this method are shown as below.

(1) Establish the total control chart of y, which denoted as t;
(2) According to the sequence of $x_n, x_{n-1}, …, x_1$, select $\{x_n\}, \{x_n, x_{n-1}\}, …, \{x_n, x_{n-1}, …, x_2\}$ as non-control cause sets, and establish the corresponding cause-selecting control charts named $t_n, t_{n-1}, …T_2$;
(3) Set $k=n$;
(4) Diagnose x_k by the following rules:
   (i) when t is out of control whereas t_k is under control, conclusion can be drawn that only x_k is abnormal cause. Adjust x_k to the normal level, and then the diagnosis is ended.
   (ii) when t and t_k are all out of control, it reveals that there are other causes except x_k are abnormal. Now it needs to consult the control chart of x_k to determine whether x_k is in normal level. If x_k is normal, go to next step, otherwise, adjust x_k to normal level firstly, and then go to next step;
(5) $K=k-1$. if $k>1$, go to previous step, otherwise, the diagnosis is ended.

in the execution of above algorithm, when comparison of control charts in one step is finished, the current cause is diagnosed. In the worst case, i.e., x_1 is an abnormal cause, the comparison times in the whole procedure is n, so the time complexity is $o(n)$, and the average is $o(n/2)$.

moreover, in every step of diagnosis, it must be confirmed that the current cause is normal, or has been adjusted to normal level, should the next step be executed. So the diagnosis is a serial procedure in which the judgment and the adjustment are executed alternately, and the period is long.

Stepwise diagnosis method

Algorithm in 2.1 has no specific requirements in the selection sequence of non-control causes. In general, the influences of different quality causes acted on the quality index are unequal. If quality causes are diagnosed by the sequence of the influences that they act on the quality index when the process is out of control, the comparison times of control charts could be reduced and the efficiency of diagnosis could be increased on the probability.

Influences of different causes are revealed by the coefficients in the regression equation. So

according to the absolute value of the coefficients, sort the causes in ascending order and then diagnose the causes by this sorted sequence, is the basic idea of stepwise diagnosis algorithm.

Influences of different causes are equal is the worst case in this algorithm, and in this case, the time complexity is $o(n)$, and the average is $o(n/2)$. When influences of causes are unequal, the average time complexity is lower than $o(n/2)$.

Theoretical basis of new algorithm

Although algorithm in 2.2 could reduce the comparison times in diagnosis on the probability, it is still a serial procedure. Therefore, it is necessary to find a new algorithm to make the judgment stage and adjustment one independent. Thus, we could find all the abnormal causes firstly, and then adjust them uniformly. In such case, the period of diagnosis could be shortened greatly and the efficiency could be increased highly.

Criteria about abnormal causes

Two criteria about abnormal causes should be introduced firstly.

criterion 1: suppose $x=\{x_1,x_2,...,x_n\}$ is the cause set of the process quality index y, $\{x_1,x_2\}$ is a set partition of x. Select x_1 and x_2 as non-control cause set respectively and establish the corresponding cause-selecting control charts about y, denoted as t_1 and t_2. when the total control chart of y is out of control, the scope of abnormal causes could be determined by the following rules:

(1) When t_1 is under control and t_2 is out of control, the abnormal causes must be contained in x_1, causes in x_2 are all in normal level, and vice versa;

(2) When t_1 and t_2 are all out of control, it can be concluded that abnormal causes are contained in both x_1 and x_2.

criterion 2: suppose process quality index y is influenced by only two causes x_1 and x_2, select x_1 and x_2 as non-control cause respectively and establish their cause-selecting control charts about y, denoted as t_1 and t_2. when the total control chart of y is out of control, abnormal causes could be confirmed by the following rules.

(1) When t_1 is under control and t_2 is out of control, x_1 is the only abnormal cause, and vice versa;

(2) When t_1 and t_2 are all out of control, both x_1 and x_2 are abnormal causes.

Criterion 2 is the particular instance of criterion 1.

Intermediate quality index

in the diagnosis procedure of algorithm in 2.1, suppose x_k is the current cause, the corresponding non-control cause set is $x_k=\{x_n,x_{n-1},...,x_k\}$, and the cause-selecting control chart is t_k. By comparing and adjusting in the previous $n-k+1$steps, it can be sure that $x_n,x_{n-1},...,x_{k+1}$ are all in normal level, so when t_k is out of control, it is obviously that x_k is an abnormal cause, and this is the reason why the whole diagnosis must be executed serial.

analyze from the aspect of mathematical model, every cause-selecting control chart in algorithm in 2.1is established on the basis of regression equations of y about the corresponding non-control cause sets, so the diagnosis result of current cause must be influenced by the states of previous ones. If the influence could be confined only between the adjacent pairs of causes in diagnosis sequence, the defect of serialization could be avoided. Thus, we can confirm all the abnormal causes firstly, and then adjust them uniformly, so the period of diagnosis could be shortened and the efficiency could be promoted. For this reason, there is the following theorem:

theorem 1: suppose y is the process quality index and influenced by the cause set $x=\{x_1,x_2,...,x_n\}$, x_1 is a subset of x. Compute the regression equation of y about x_1, denoted as $y_1=f(x_1)$, and select x_1 as non-control cause set to establish the cause-selecting control chart about y, denoted as t_1. further

assume that x_2 is a subset of x_1, $y_2=g(x_2)$ is the regression equation of y_1 about x_2. select x_2 as non-control cause set to establish the cause-selecting control chat about y_1, denoted as t_2, the following conclusion can be founded:

(1) If and only if abnormal causes are contained in x-x_1, should t_1 be out of control;

(2) If and only if abnormal causes are contained in x_1-x_2, should t_2 be out of control;

(3) The statuses of causes in x-x_1 have no relationship with the state of t_2.

conclusion (1) and (2) could be proved by criterion 1 in 3.1. for the last conclusion, considering the regression equation $y_1=f(x_1)$, we know that y_1 is the function of x_1, and only influenced by the elements in set x_1, so the quality causes in x-x_1 has no effect on y_1, and naturally has no relationship with t_2.

from theorem 1 it can be concluded that in the establishment of multistage cause-selecting control charts, if the dependent variable in the regression equation of previous stage is chose as new quality index to establish corresponding cause-selecting control chart of non-control cause set in current stage, then the state of cause-selecting control chart of new quality index is only affected by the cause set which selected as the independent variables in the regression equation of new quality index, and has no relationship with the other causes.

new quality indices calculated by regression equations in different stages are called as intermediate quality indices. The interaction of different quality causes can be confined only between the adjacent pairs in the diagnosis sequence by using the intermediate quality indices, so the defect of serialization in previous algorithm could be avoided, and the period of diagnosis could be shortened greatly.

Algorithm based on binary tree and intermediate quality indices

the basic idea of new algorithm is using the method of bisection to divide the quality cause set progressively, until every subset contains only one element. The result of this division is a complete binary tree of quality causes. Then for every node, select the corresponding subset as non-control cause set to establish cause-selecting control chart of the intermediate index in its parent node. In diagnosis, start from the root node and use the criterion 1 repeatedly to narrow the scope of abnormal causes until reach to the leaf node, and then criterion 2 could be used to confirm the abnormal quality causes.

Division of quality causes

suppose y is the process quality index and influenced by quality causes $x_1,x_2,...,x_n$, let $n_1=[n/2]$, where operator $[. \ ]$ Means integer conversion. After the first division, set $x=\{x_1,x_2,...,x_n\}$ is divided into two subsets x_{11} and x_{12}, where $x_{11}=\{x_1,x_2,...,x_{n1}\}$ and $x_{12}=\{x_{n1+1},x_{n1+2},...,x_n\}$.

if elements in set x_{11} is more than 1, set $n_2=[n_1/2]$, and divide x_{11} into two subsets $x_{21}=\{x_1,x_2,...,x_{n2}\}$ and $x_{22}=\{x_{n2+1},x_{n2+2},...,x_{n1}\}$. In the same way, if the number of elements in set x_{12} is greater than 1, set $n_3=[(n_1+n)/2]$, and x_{12} is divided into two subsets x_{23} and x_{24}, where $x_{23}=\{x_{n1+1},x_{n1+2},...,x_{n3}\}$, $x_{24}=\{x_{n3+1},x_{n3+2},...,x_n\}$. In the second division, the amount of subsets is not more than 4.

further division is continued on the subsets which obtained in the second division. If a subset contains more than 1 element, it must be divided in the same way. As a general, in the kth division, the number of subsets is 2^k at most, and these subsets are denoted as $x_{k1},x_{k2},...,x_{km}$, where $m \leqslant 2^k$.

this division should be repeated until every subset contains only one element, then the operation is in the end.

the whole steps of division could be shown as a complete binary tree in figure 1.
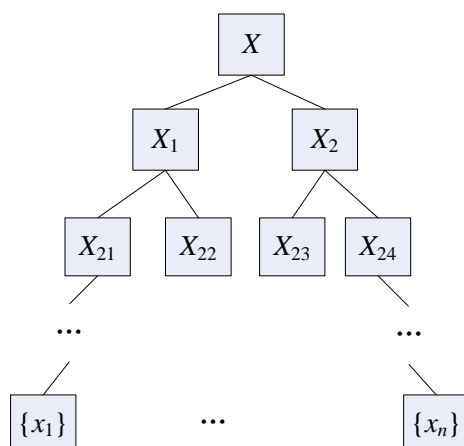
Figure 1. complete binary of quality causes set

Establishment of control charts

according to theorem 1 and the division result of quality cause set shown in figure 1, establish control charts by the following principles:

(1) For the root node and the corresponding set x, select no element as non-control cause to establish the total control chart t of the process quality index y.

(2) For the first level, the intermediate nodes are x_{11} and x_{12}. calculate the regression equation of y about x_{11}, i.e., $y_{11}=f_{11}(x_{11})$, and another regression equation of y about x_{12}, which denoted as $y_{12}=f_{12}(x_{12})$. Then based on the regression equations, select x_{11} and x_{12} as non-control set respectively, establish the corresponding cause-selecting control charts of y, which denoted as t_{11} and t_{12}.

(3) As a general, for any node x_{kj} in the kth level, assume its parent node is $x_{k-1,i}$, calculate the regression equation of intermediate quality index $y_{k-1,i}$ about x_{kj}, i.e., $y_{k-1,i}=f_{kj}(x_{kj})$, and then select x_{kj} as non-control cause set to establish cause-selecting control chart t_{kj} of intermediate quality index $y_{k-1,i}$.

the intermediate quality indices and control charts form another tree with the same structure shown in figure 1. the quality index, non-control cause set and chart in the same location could be drawn in one node, and the result could be shown as figure 2. in this figure, the non-control cause set in root is empty, and the chart t is total control chart of y. In the other nodes, x_{ij} is the non-control cause set, and t_{ij} is the cause-selecting control chart of the intermediate quality index in its parent node.
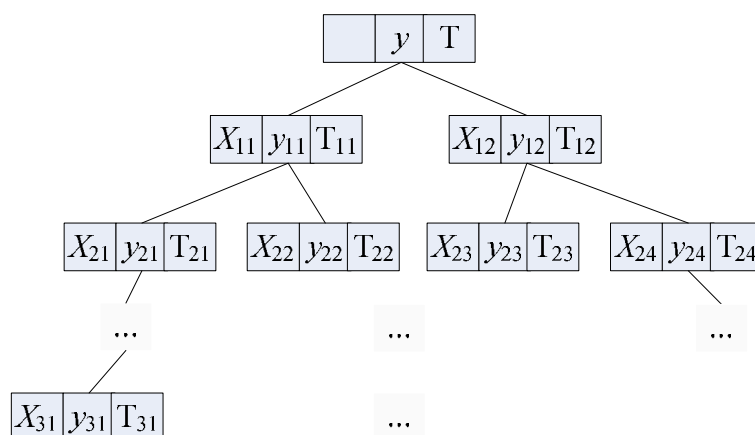

Figure 2. complete binary tree of non-control cause set, intermediate quality and control chart

Diagnosis procedure

according to the structure shown in figure 2, the abnormal quality causes could be diagnosed by the following principles:

(1) If t is normal, it means that the process is under control, and the diagnosis is ended;

(2) If t is out of control, in the first level of the tree at least one chart of t_{11} and t_{12} is out of control. Now operation should be done as below:

(i) if t_{11} is under control and t_{12} is out of control, conclusion can be drawn that the abnormal quality causes must be contained in x_{11}, and causes contained in x_{12} are all in normal level. Further operation only should be done on the left sub tree in order to find the abnormal quality causes;

(ii) if t_{11} is out of control and t_{12} is under control, it reveals that the abnormal causes must be contained only in x_{12}, so the further operation should be done on the right sub tree;

(iii) when t_{11} and t_{12} are all out of control, it means that the abnormal causes are contained both in x_{11} and x_{12}. now further operation should be done on the left sub tree and the right one.

(3) In general, for the sub tree whose root node is x_{ij}, suppose its child nodes are $x_{i+1,k}$ and $x_{i+1,k+1}$, the operation done on the sub tree is as follow:

(i) if $t_{i+1,k}$ is under control and $t_{i+1,k+1}$ is out of control, it means that the abnormal causes are contained in $x_{i+1,k}$. If $x_{i+1,k}$ only contains one element, this element must be an abnormal cause, otherwise the same operation should be done on the sub tree whose root node is $x_{i+1,k}$.

(ii) if $t_{i+1,k}$ is out of control and $t_{i+1,k+1}$ is under control, it means that the abnormal causes are contained in $x_{i+1,k+1}$. in the same way, if $x_{i+1,k+1}$ contains one element, then this one is an abnormal cause, otherwise the same operation should be done on the sub tree whose root node is $x_{i+1,k+1}$.

(iii) when $t_{i+1,k}$ and $t_{i+1,k+1}$ are all out of control, we can conclude that both $t_{i+1,k}$ and $t_{i+1,k+1}$ contain abnormal causes. If $x_{i+1,k}$ contains one element, then this element is in abnormal level, otherwise it need to diagnose the sub tree which rooted as $x_{i+1,k}$. In the same way, if $x_{i+1,k+1}$ contains one element, this one is in abnormal level, otherwise the same operation should be done on the sub tree rooted as $x_{i+1,k+1}$.

(4) Repeat step (3) until all the abnormal causes are found.

the above algorithm could diagnose all the abnormal causes continuously and make the judgment procedure independent to the adjustment one, so the defect of serialization in algorithm in 2.1 could be avoided, and the period is shortened greatly and the efficiency is increased highly. Further more, according to the characteristics of complete binary tree, in the worst case, i.e., the current quality cause is a leaf node in the last level, the comparison times of control charts is $[log_2n]$. Comparing to the time complexity of $o(n)$ in algorithm in 2.1, efficiency of the new algorithm is promoted obviously.

Example of application

In one process, the quality index y is influenced by 6 quality causes, denoted as $x=\{x_1, x_2, ..., x_6\}$. In the steady state of process, establish the complete binary tree of the quality cause set division, the intermediate quality indices and the corresponding charts as shown in figure 3.
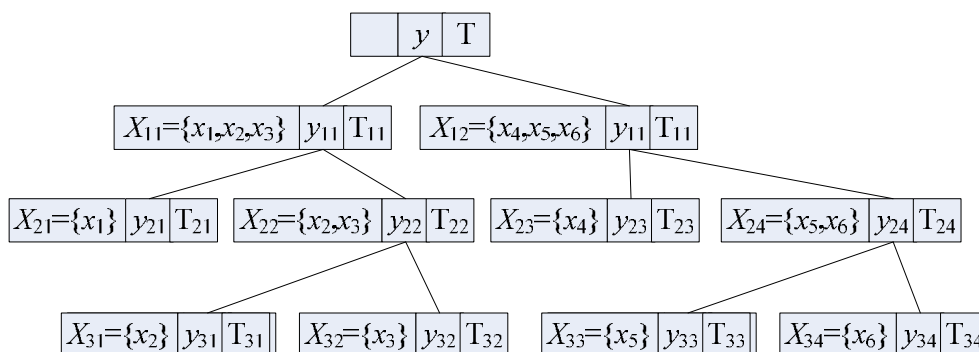
Figure 3. complete binary of example

At one moment, the control chart t is out of control, and the cause-selecting control charts t_{12}, t_{21}, t_{22} and t_{31} are both out of control at the same time, the diagnosis of abnormal causes is as follow:

(1) In the first level, only t_{12} is out of control, so the abnormal causes are only contained in x_{11}. now the left sub tree rooted as x_{11} should be handled in the next step;

(2) In the second level, t_{21} and t_{22} are both out of control, so both x_{21} and x_{22} contain abnormal causes. In set x_{21}, there is only one element x_1, so x_1 is an abnormal cause. The elements in x_{22} is more than 1, so further operation should be done on the sub tree rooted as x_{22};

(3) In the last level, t_{31} is out of control and t_{32} is under control, it reveals that only x_{32} contains abnormal causes. Now x_{32} only has one element, so x_3 is another abnormal cause;

(4) Now the diagnosis is finished, and conclusion can be drawn that x_1 and x_3 are the abnormal causes.

Conclusion

Aiming at the multi-causes process quality diagnosis, this paper proposed an algorithm based on the progressive division of quality cause set and the concept of intermediate quality index. The complete binary tree corresponding to the division of quality cause set could reduce the comparison times of control charts efficiently, and the introduction of intermediate quality index could confine the influences of different causes only between the pairs which adjacent in the diagnosis sequence, and make the judgment procedure independent to the adjustment one, so the period is shortened and the efficiency is increased.

Acknowledgements

Reference

[1] pham.d.t and waini m. A. Feature-based control chart pattern recognition. International journal of production research. Vol.35(1997), p.1875-1890.

[2] zhonghua yu, and zhaotong wu. Algorithm of quality control in small batch manufacturing process. Journal of mechanical engineering. Vol.37(2001), p.60-64.

[3] rui miao, xiaoming sun, and shugang li, et al. Research on statistical process quality control based on low volume manufacturing. Computer integrated manufacturing system. Vol.11(2005), p.1633-1635

[4] gongxu zhang. A new diagnosis theory with two kinds of quality. Total quality management.

Vol.1(1990) , p.249-257.

[5] huiying zheng. Multi-cause stepwise diagnosis with two kinds of quality. Journal of university of science and technology beijing. Vol.18(1996), p.9-12.

[6] daoyu liu, pingyu jiang. E-quality control architecture for multistage machining processes. Computer integrated manufacturing systems. Vol.13(2006), p.782-790.

[7] hyun, kwang and myong. Online monitoring and diagnosis of batch processes: empirical model-based framework and a case study. International journal of production research. Vol.44(2006), p.2361-2378.

[8] shaoxiong wu. Intelligent statistical process control system. Computer integrated manufacturing system. Vol.12(2006), p.1832-1837

[9] haijing tian, and erdun pan. Multivariate statistical process control for auto correlated process. Journal of shanghai jiaotong university. Vol.42(2008), p.496-499.

[10]xingyu jiang, shijie wang, and kai zhao, et al. Intelligent process quality control system for networked manufacturing. Journal of mechanical engineering. Vol.46(2010), p.186-194

[11]a.shaban，m.shalaby，and e.abdelhafiez，et al. 2010. automated identification of basic control charts patterns using neural networks. Journa1 of software engineering and app1ications. Vol.3(2010), p.208-220

[12]r.s.guh. Simultaneous process mean and variance monitoring using artificial neural networks. Computers and industrial engineering. Vol.58(2010), p.739-753

[13]feng jia, pingyu jiang, daoyu liu, et al. Error propagation control method for multistage batches machining processes of blades. Computer integrated manufacturing system. Vol.18(2012), p.76-88

[14]jin sun, liu yinhua and lin zhongqin. A bayesian network approach for fixture fault diagnosis in launch of the assembly process. International journal of production research. Vol.50(2012), p.6655-6666