A Classification Techniques For Quality Improvement*

Xu Jichao Liu Yumin Zhang Li

Zhengzhou Insitute of Aeronautics

Zhengzhou, Henan, 450052, China

Email: jxu@zzia.edu.cn

Abstract

As we know, the quality of processes is technically depicted by variation, a product or process with the best quality must naturally require the variation as less as possible. The variation is usually reduced with many ways, say, by adjusting parameters settings under robust design with many turns expensive experiments. So ones are trying to reach the robustness by detecting cheap and simple methods. In this paper, a both practical and simple technique for quality improvement, namely reducing the variation, by data classification is studied. First, all possible system factors are included, which may dominate the variation law. And then we make use of the past observations and their classification as well as boxplot charts to find out the internal rule between the variation and the system factor. Next, adjust the location of the system factor according to the rule so that the variation could, to some extent, be lessened. Finally, two typical quality improvement cases based on data classification are presented.

Key words : Robust Design, Data Classification, Boxplot Chart, Variation Reduction

1. Introduction

The quality of products or processes is to be technically depicted by using the magnitude of the variation around target of interesting quality characteristic. In order to improve the quality, ones have been working on variation reduction. Surely there exist several ways to reduce the variation. The easy one is to change new equipment and to purchase advanced materials. This is not what quality engineering is for because of the high cost with this way. The main viewpoint of quality professional for quality

* Supported by National Science Funds for Distinguished Young Scholars 70125004, and NSFC Projects 9870100,700072029

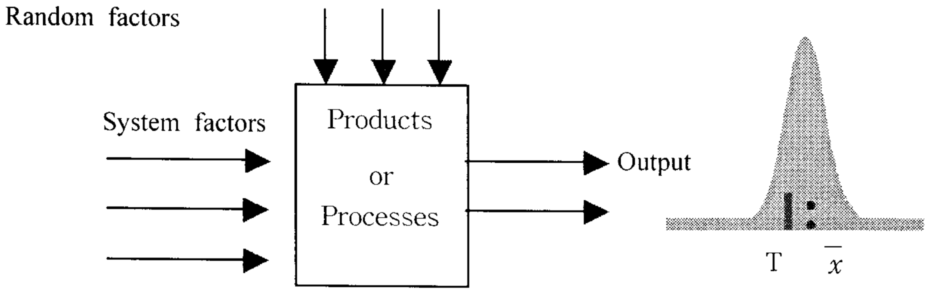

Fig.1 The general model of robust design

improvement is to find out cheap way, Usually, the parameter design technique referred to Dr. Taguchi is employed for the above objective. Thousands of cases study have proved that the robust parameter design is very effective. The general model of robust design studied by parameter design method can be described as follows (See Fig.1):

The parameter design desensitizes effect of the noise on system factors by optimizing settings of system factors. So the variation around the target of interesting quality characteristic can be reduced.

Although the parameter design methods are widespread considered as an attractive optimization technique for quality improvement, there still exist many disadvantages such as:

1. This off-line experiment often interrupts production
2. The operation and planning for experiment are relatively complicated
3. The cost with the experiment usually is comparatively expensive in contrast with on-line quality improvement.

In real production process, many historical data about the process have been observed, the data are very informative for quality improvement of the process. Key problems are how to make use of the data. In this paper, we present a technique for quality improvement only by historical data classification, which can cheaply and easily be implemented. The second section of this paper reports an identification method of sensitive system factor for reducing variation of the process. In the third section, the two methods of data classification related to continuous and categorical system (or controllable) factors are respectively presented, furthermore, operating approach for the variation reduction are given. Two cases indicated implementation procedures of the data classification technique are employed in the fourth section. Finally, we conclude the paper with several comments on the data classification methods.

2. Identification Approach for Significant System Factors

For convenience, we denote the interesting quality characteristic by Y, and system factor by a vector $X=(X_1, X_2, \ldots X_k)$, in which system factors, namely components of X, are often divided into two types, continuous variable type such as strength, temperature etc., and categorical type such as operator, tool, machine etc. In order to monitor quality characteristic of process or products, the n observations for the quality characteristic Y, are regularly recorded as $y_1, y_2, \ldots y_n$. It is by the data that we are able to evaluate and find out the functions of various system factors, $X_1, X_2, \ldots X_k$, further improve the quality of process by reducing variation of Y.

In this section, we firstly present how to discriminate the significant system factor or a group of system factors for improving process. The system factors, $X_1, X_2, \ldots X_k$, may be interacted closely, may be not. Suppose firstly that all of system factors $X_1, X_2, \ldots X_k$ are not interacted, single factor, X_i, need to be checked with one by one. The following methods are used to show the basis of discrimination. The exact relationship between quality characteristic Y and various factors can be written as follows

$$Y = f(\underbrace{X_1, X_2, \ldots X_k}_{\text{System factors}},\ \underbrace{R_1, R_2, \ldots}_{\text{Random factors}})$$

In order to identify the significant system factor, the relation can be simplified as

$$Y = g(X_i) + \varepsilon(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n, R_1, R_2, \cdots) \qquad (1)$$

Where ε indicate various random error caused by the fluctuation of $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots X_n, R_1, R_2, \ldots$. Although system factors usually are controllable and their settings are determined in advance, the variation around the settings cannot be avoided absolutely. For example, in categorical system factor, say, operator's system factor, the setting of operator is that each operator i ($i=1,2,\ldots,m$) should be qualified for operating the current process. The difference, however, among operators cannot be avoided, so the random error caused by system factor is then generated. In the similar way, the continuous system factor can also result in random error, for instance, the setting of temperature only is an expectation value, namely the mean of the temperature. In practice, the temperature has to fluctuate around the expectation. So the random error, which is from the fluctuation, is also followed. Of course, the random factor R is the main root source of

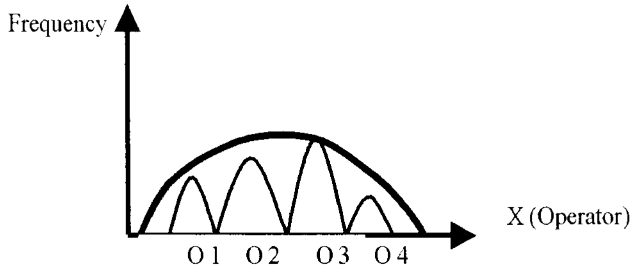

Fig. 2 Categorical system factor

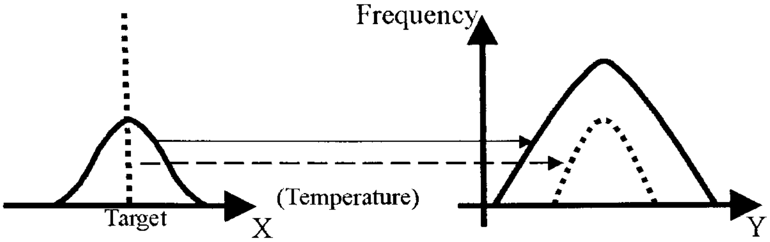

Fig. 3 Continuous system factor

random error. Usually, ε is often supposed to be of some kind of distribution, but we are not going to discuss the exact distribution here. According to the above analysis, the relation between the *i*th system factor X_i and Y can be illustrated by the following graph. As shown Fig.2 and Fig.3.

Firstly, we discuss the recognition of significant system factor for categorical type. Without loss of generality, the typical example, operator={operator1, operator2, ... , operator *m*}, is employed to indicate the procedure of recognition. Suppose that $y_1, y_2, \ldots, y_n$ are the n observations that all operators possess during their shifts.

Step 1 Group the data or observations according to operators.

operator 1 (x_1) $y_{11}, y_{12}, \ldots, y_{1k1}$;

operator 2 (x_2) $y_{21}, y_{22}, \ldots, y_{2k2}$;

operator m (x_m) $y_{m1}, y_{m2}, \ldots, y_{mkm}$;

where $\sum_{i=1}^{m} k_i = n$

Step 2 Compute the (S/N), ratio of each operator *I*

$$\left(\frac{S}{N}\right)_i = \left(\frac{\mu_i^2}{\sigma_i^2}\right), \quad (i=1,2,\cdots,m)$$

The estimations of μ_i and σ^2 are

$$\hat{\mu}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} y_{ij};$$

$$\hat{\sigma}_i^2 = \frac{1}{k_i - 1}\sum_{j=1}^{k_i}(y_{ij} - \hat{\mu}_i)^2$$

Step 3 Compute the spread or variance of $(S/N)_i$ ratio for operator l.

Note that the variance of (S/N)i can be used to measure whether or not the operato r's system factor is relatively significant. Similarly we also get the result of continuous type.

For the continuous system factor, say, X, which is quality characteristic of proceeding process and affects the Y. Usually we also have the m observation $x_1, x_2, \ldots, x_m$ for X. Simulate the situation of categorical system factor type, the continuous type can be transferred into categorical type, and then deal with the categorical type as above.

Firstly, establish histogram of X as shown Fig.4, in which $a_1 = \min\{x_1, x_2, \ldots, x_m\}$; $a_l = \max\{x_1, x_2, \ldots, x_m\}$;. $[a_i, a_{i+1})$ is the ith interval, $i = 1, 2, \ldots, l-1$. According to each interval, we classify the data of Y. $y_1, y_2, \ldots, y_n$, into l-1 groups of the data. The groups method for $y_1, y_2, \ldots, y_n$ is as follows. The group consisted of $y_{i1}, y_{i2}, \ldots, y_{iki}$ on $[a_i, a_{i+1})$ means all observations of Y in which the observations of system factor belong to interval $[a_i, a_{i+1})$ So discrimination method for significant continuous system factor can be summarized as:

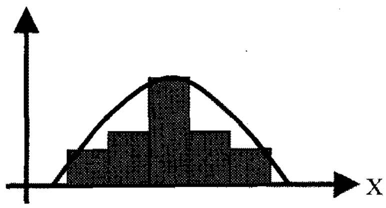

Fig. 4 Histogram of X

Step 1 Group the observations of Y according to the intervals.

$$\begin{array}{lllll} [a_1, a_2): & y_{11}, & y_{12}, & \Lambda, & y_{1k_1} \\ [a_2, a_3): & y_{21}, & y_{22}, & \Lambda, & y_{2k_2} \\ \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\ [a_{l-1}, a_l): & y_{l-11}, & y_{l-12}, & \Lambda & y_{l-1k_{l-1}} \end{array}$$

Step 2 Computer the $(S/N)_i$ ratio of each interval.

$$(S/N)_i = \mu_i^2 / \sigma_i^2, \quad (i = 1, 2, \ldots, l-1)$$

where the μ_i and σ_2^i have the same meaning as the above we defined in categorical situation.

Step 3 Compute the spread or variance of $(S/N)_i$ of X.

The variance is used to indicate the significant degree of system factor X.

In general, the criteria for judging significant degree are that the larger variance of system factor is, and then we make use of the significant factor to reduce the variation of process output.

3. The Variation Reduction by Grouping of Observations

As above what we have shown, continuous system factors are of very similar expression with the categorical by introducing interval as describing points. Here we will further define that mid-point, $(a_i + a_{i+1})/2$, of each interval $[a_i, a_{i+1})$ is used to represent the interval. In this way, we will have the exact same method as the categorical to deal with quality improvement problems, so hereafter, we only present the method of reducing variation for the categorical situation.

Suppose that X is both significant and easily operational, the set consisted of possible values of X is denoted by $X = x_1, x_2, \ldots, x_m$. xi may represent operator i, tool i, machine i or mid-point $(a_i + a_{i+1})$ of the interval under continuous system factor situation. Based on $x_1, x_2, \ldots, x_m$, the groups of observations of Y have been already obtained such as

$$\begin{array}{lllll} x_1: & y_{11}, & y_{12}, & \Lambda & y_{1k_1} \\ x_2: & y_{21}, & y_{22}, & \Lambda & y_{2k_2} \\ \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\ x_m: & y_{m1}, & y_{m2}, & \Lambda & y_{mk_m} \end{array}$$

The problem that we are dealing with is to search for a X_{i0} which reach the largest of $(S/N)_i$ ratio. The direct method is to calculate the $(S/N)_i$ for each group of observations and then compare with each other. The X_{i0} which reaches the largest of $(S/N)_i$ is naturally selected to be an optimal level. If the significant system factor that we are discussing is operator, then we should adjust the each operator's operational level to one of operator i_0, namely optimal level. If tool, all of the tools should be changed to the standard tool i_0, optimal tool i_0 selected. If it is the quality characteristic of preceding process, X, we should adjust old average of the preceding process to x_{i_0} in which

$$x_{i_0} = \frac{a_{i_0} + a_{i_0+1}}{2}$$

a_{i_0}, a_{i_0+1} are end-points of interval $[a_{i_0}, a_{i_0+1})$, and so on.

For both simply and straight, we introduce the boxplot chart to find out optimal level, x_{i_0}, as shown Fig.5.

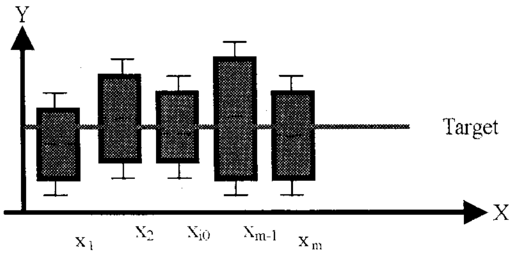

Fig. 5 Boxplot chart

Where the box include 50% of observations under x_i, the dotted line on the middle of box represents the average of all observations under x_i. So the boxplot chart is used to find out which x_i is optimal point. According to Fig.5, x_{i_0} is the optimal setting what we are wanting.

In without interaction effect, we can respectively investigate each significant system factor to obtain optimal levels is new settings we are wanting, which will greatly improve the quality, that is, the variation of Y will to great extent be reduced. Interaction will effect on reproducibility, ones often try to avoid the interaction by selecting quality characteristic such as S/N of Taguchi's or generic function. In reality, however, it is an extremely difficult problem to choose an available quality characteristic not involved in interaction. As we know, an ideal function is not easily to be got, S/N ratio can not guarantee to avoid interaction. If interaction effected the investigated optima level of each system factor will not be available to obtain an integral optimal setting combination. In this case, we must select the significant system factor group in which the system factors are interacted heavily. For example, we suppose two system factors X and X^* are heavily interacted, that is

$$X=\{x_1, x_2, \ldots, x_{m1}\}$$

$$X^*=\{x_1^*, x_2^*, \ldots, x_{m2*}\}$$

The suggestion will be to arrange for possible combination of Xand X^*, namely

$$X \times X^*=\{x_1x_1^*, \ldots, x_1x_{m2}*; \ldots\ldots; x_{m1}x_1^*, \ldots, x_{m1}x_{m2}*\}$$

In this way, the grouping of observations of Y is changed as follows.

$$x_ix_j^*:\ y_{i_1}, y_{i_2}, \Lambda, y_{ik_i}, y_{j_1}^*, y_{j_2}^*, \Lambda, y_{jk_j}^*,$$

$$(i=1,2,\ldots,m1\ ;\ j=1,2,\ldots,m2)$$

Here $y_{j_1}^*, \Lambda, y_{jk_j}^*$, is the group of observations of Yunder x_j^*. The optimal level or significant system factor group will be found out in $X \times X^*$. The rest of what we will do for recognition of significant system factor group and reducing variation study are the exactly same as the above what we have presented in terms of univariate. Thus these contents will not be included here.

4. Cases Study for Quality Improvement by Grouping of Data

Case 1

This case is involved in two categorical significant system factors. One is operators consisted of operator 1, operator 2, and the other is tools, consisted of tool 1, tool 2, and they are not interacted. So we will respectively investigate optimal levels for reducing variation. The interesting quality characteristic of the process is Y with target value 2.5, and historical observations for Y were exactly recorded as

1.98, 2.65, 2.50, 2.80, 1.90, 3.00, 2.76, 1.96, 2.96, 3.60
1.87, 2.67, 1.67, 2.10, 3.50, 2.02, 2.99, 1.98, 2.18, 1.98

The grouping under operators:
Operator 1:
1.98, 2.65, 2.50, 2.80, 1.90, 1.87, 2.67, 1.67, 2.10, 3.50

Operator 2:
3.00, 2.76, 1.96, 2.96, 3.60, 2.02, 2.99, 1.98, 2.18, 1.98

The grouping under tools:
Tool 1:
1.98, 2.65, 2.50, 2.80, 1.90, 3.00, 2.76, 1.96, 2.96, 3.60
Tool 2:
1.87, 2.67, 1.67, 2.10, 3.50, 2.02, 2.99, 1.98, 2.18, 1.98

The boxplot chart for operators as shown Fig.6.

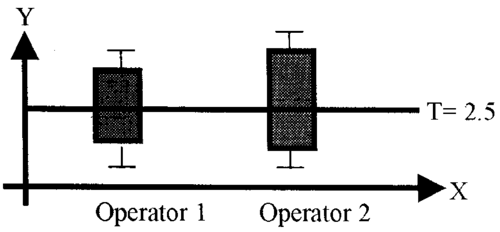

Fig. 6 The boxplot chart for operators

So, the operator 2 is optimal level.

The boxplot chart for tools as shown Fig.7.

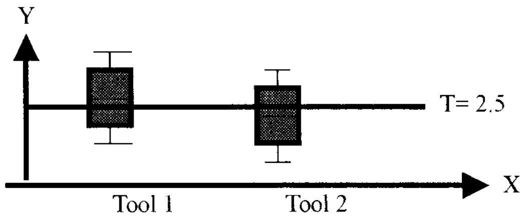

Fig. 7 The boxplot chart for tools

From Fig.7, tool 1 is the best level.

Due to the consideration without interaction, (operator 2, tool 1) is an optimal level combination. Thus we train another operator toward operator 2 and adjust tool 2 to hit tool 1 level, the original variation will be greatly reduced.

Case 2

In this case, a continuous significant system factor, affecting output Y of the process, X, as a quality characteristic of preceding process is investigated.

The observations of X were recorded as

14.6, 10.6, 11.0, 12.3, 10.8, 9.60, 8.90, 13.5, 14.1,
9.90, 9.89, 11.2, 13.1, 12.5, 14.1, 9.70, 10.9, 9.50,
12.6, 12.5, 15.0, 9.10, 11.4, 11.2, 16.1, 10.7, 12.2,
8.90, 15.6, 14.1, 10.6, 11.5, 11.4, 9.70, 9.90, 10.5,
11.6, 12.3, 10.5, 14.6, 8.70, 15.1, 14.6, 8.90, 9.20,
11.6, 12.5, 11.4, 10.9, 12.3, 11.4, 10.5, 12.6, 11.6,

The histogram of observations of X made is as Fig.8.

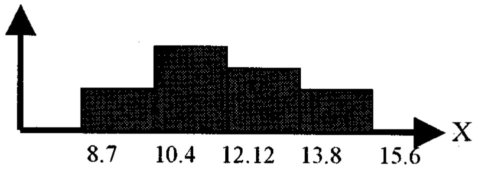

Fig. 8 The histogram of X

In the above histogram, the range of the observations is divided into four intervals, [8.7,10.4), [10.4,12.12), [12.12,13.8), [13.8,15.6). The midpoints are 9.55, 11.26, 12.96 and 14.7 respectively. According to the data felled in the intervals and corresponding the observations of Y are grouped as follows.

15.1, 14.6, 17.6, 16.1, 12.3, 15.0

mid-point: 9.55
10.6, 11.5, 14.6, 15.9, 16.8, 17.6
13.5, 14.7, 16.1, 17.5, 18.1, 10.1
12.1, 13.6, 17.1, 18.1, 14.1, 15.1
9.50, 15.5, 14.6, 17.1, 13.2, 16.1

mid-point: 11.26
11.7, 13.6, 12.3, 16.1, 17.0, 15.6
15.2, 16.3, 14.9, 14.1, 16.1, 13.9
17.0, 13.9, 12.8, 15.1, 15.0, 14.9
14.4, 16.7, 15.8, 16.0, 12.9, 16.1
14.6, 15.1, 16.1, 17.0, 14.8, 15.1

mid-point: 12.69
15.1, 16.0, 14.8, 16.0, 13.5, 14.2
14.2, 18.0, 12.9, 16.0, 15.1, 15.6
12.4, 14.7, 14.8, 14.0, 15.0, 15.6
14.5, 15.0, 15.6, 15.8, 16.1, 17.0

mid-point: 14.7
12.9, 14.8, 15.2, 16.0, 15.3, 14.6
14.1, 15.2, 14.8, 15.0, 13.9, 14.9
15.6, 16.8, 12.9, 14.8, 14.1, 15.1
14.6, 13.9, 15.0, 13.8, 14.5, 15.1

Drawing boxplot chart for each group as shown Fig.9.

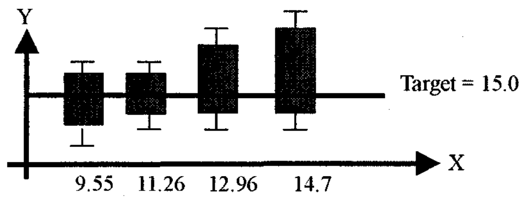

Fig. 9 The boxplot chart for groups

So, 11.26 is taken as an optimal level of X.

5. Conclusion

The quality improvement method given in this paper is of many advantages.

1. The historical observations, which are often wasted, are effectively used.
2. The observations on-line is much more cheaper and easier than experiment.
3. Implementation for classification technique of data is easy way.
4. The results of the quality improvement,

namely reducing variation, are often very effective and significant.

When there exist interaction among significant system factors, however, identification of the significant system factor group with closed interaction is quite difficult. Even if the group has been found out, if each system factor in the group takes many values of level, say, the number of members of group is equal to ten and each member takes four values of level, the number of groups for the data should be 4^{10}. The amount of computation is relatively overwhelming. In this case, we have to select some alternatives from all alternatives of 4^{10} as representatives by an available arrangement with orthogonal array.

References

1. Vijayan, N. Nair (1992), Taguchi's Parameter Design: A Paned Discussion, Technometrics, Vol. 34 (2), 127.
2. Taguchi, G. (1959), Linear Graphs for Orthogonal Arrays and Their Application to Experimental Design, With the Aid of Various Techniques, Reports of Statistical Application Research, JUSE, 6, 1-43.
3. Taguchi, G.(1987), System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost, White Plains. NY: UNIPUB/Kraus International.
4. Box, G. E. P. (1988), Signal-to-Noise Ratios, Performance Criteria and Transformations (with discussion), Tecnometrics, Vol.30, 1-40.
5. Miller, A. and Wu, C.F.J. (1991), Improving a Calibration System through Designed Experiments Technical Report, University of Waterloo, IIQP.