Part III

Industrial Engineering Applications

Fuzzy Process Control with Intelligent Data Mining

Murat Gülbay and Cengiz Kahraman

Department of Industrial Engineering, Istanbul Technical University, Macka 34367, Istanbul, Turkey
gulbaym@itu.edu.tr, kahramanc@itu.edu.tr

Abstract. The quality-related characteristics cannot sometimes be represented in numerical form, such as characteristics for appearance, softness, color, etc. In this case fuzzy set theory can handle this problem. This chapter develops fuzzy control charts for linguistic data. Later, unnatural pattern analyses are made using the probability of a fuzzy event. Unnaturalness of the linguistic data is searched with an intelligent data mining procedure.

1 Introduction

The boundaries of classical sets are required to be drawn precisely and, therefore, set membership is determined with complete certainty. An individual is either definitely a member of the set or definitely not a member of it. This sharp distinction is also reflected in classical process control charts, where each process is treated as either "in control" or "out of control". However, most sets and propositions are not so neatly characterized. It is not surprising that uncertainty exists in the human world. To survive in our world, we are engaged in making decisions, managing and analyzing information, as well as predicting future events. All of these activities utilize information that is available and help us try to cope with information that is not. Lack of information, of course, produces uncertainty, which is the condition where the possibility of error exists. Research that attempts to model uncertainty into decision analysis is done basically through probability theory and/or fuzzy set theory. The former represents the stochastic nature of decision analysis while the latter captures the subjectivity of human behavior. When the data used to construct process control charts are incomplete, vague, or linguistic, classical process control charts fail to determine the nature of the process. Therefore, a fuzzy approach to process control charts are necessary to adopt.

Fuzzy sets were introduced in 1965 by Lotfi Zadeh with a view to reconcile mathematical modeling and human knowledge in the engineering sciences. Since then, a considerable body of literature has blossomed around the concept

Murat Gülbay and Cengiz Kahraman: *Fuzzy Process Control with Intelligent Data Mining*, Studies in Computational Intelligence (SCI) **5**, 315–336 (2005)
www.springerlink.com

of fuzzy sets in an incredible wide range of areas, from mathematics and logics to traditional and advanced engineering methodologies. Applications are found in many contexts, from medicine to finance, from human factors to consumer products, from vehicle control to computational linguistics, and so on ... Fuzzy logic is now currently used in the industrial practice of advanced information technology.

Basically, when a point on the control chart is falling outside of the three-sigma control limits it shows an out of control situation. There are some interesting questions related to the computation of the probability that a chart will be out of control even all points on the chart are within three-sigma limits. Based on the expected percentages in each zone, sensitive run tests can be developed for analyzing the patterns of variation in the various zones on the control chart. The concept of dealing with probability of the data pattern on the control chart is known as "unnatural pattern analysis". Whenever a point is drawn on the control chart, the rules of accepting a pattern as unnatural should be examined. Analysis of unnatural patterns can be discovered through intelligent data mining.

Data mining (DM) is a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data. The main data mining application areas are marketing, banking, retailing and sales, manufacturing and production, brokerage and securities trading, insurance, computer hardware and software, government and defense, airlines, health care, broadcasting, and law enforcement.

Intelligent data mining (IDM) is to use intelligent search to discover information within data warehouses that queries and reports cannot effectively reveal and to find patterns in the data and infer rules from them, and to use patterns and rules to guide decision-making and forecasting. Main tools used in intelligent data mining are case-based reasoning, neural computing, intelligent agents, and other tools (decision trees, rule induction, data visualization).

In recent years, the need to extract knowledge automatically from very large databases has grown. In response, the closely related fields of knowledge discovery in databases (KDD) and data mining have developed processes and algorithms that attempt to intelligently extract interesting and useful information from vast amounts of raw data. The term DM is frequently used to designate the process of extracting useful information from large databases. The term KDD is used to denote the process of extracting useful knowledge from large data sets. DM, by contrast, refers to one particular step in this process. Specifically, the data mining step applies so-called data mining techniques to extract patterns from the data. Additionally, it is preceded and followed by other KDD steps, which ensure that the extracted patterns actually correspond to useful knowledge. Indeed, without these additional KDD steps, there is a high risk of finding meaningless or uninteresting patterns [4, 8]. In other words, the KDD process uses data mining techniques along with any required pre- and post-processing to extract high-level knowledge from low-level data. In practice, the KDD process is interactive and iterative, involving numerous steps with many decisions being made by the user. DM techniques are essentially pattern discovery algorithms. Some techniques such as association rules are unique to data mining, but most are drawn from related fields such as databases, statistics, pattern recognition, machine learning, neurocomputing, and artificial intelligence.

The application of data mining to fuzzy process control has not yet been studied extensively. This chapter is organized as follows. In Sect. 2, the basics of process control charts are summarized. Next, fuzzy process control charts are explained in Sect. 3. Fuzzy Process Control with Intelligent Data Mining is developed in Sect. 4. Finally a numerical application in Sect. 5, and conclusions in Sect. 6 are given.

2 Process Control Charts

Based on the statistical variation of any process – control charts help focus on stability of a process. An essential element of producing a high quality product is insuring that the characteristics of that product remain constant over time. Statistical process control charts are widely used to determine whether a process is stable and to monitor that process over time. When the characteristics of interest can be measured (rather than simply observed), it is common to take periodic samples of measurements and then plot statistic such as the mean and range. It is used to determine how much variability in a process is due to random variation and how much is due to unique events/individual actions so that you know whether or not the process is in statistical control. A typical control chart plots the sample statistics together with upper and lower control limits shifted from center line (CL) as shown in Fig. 1.

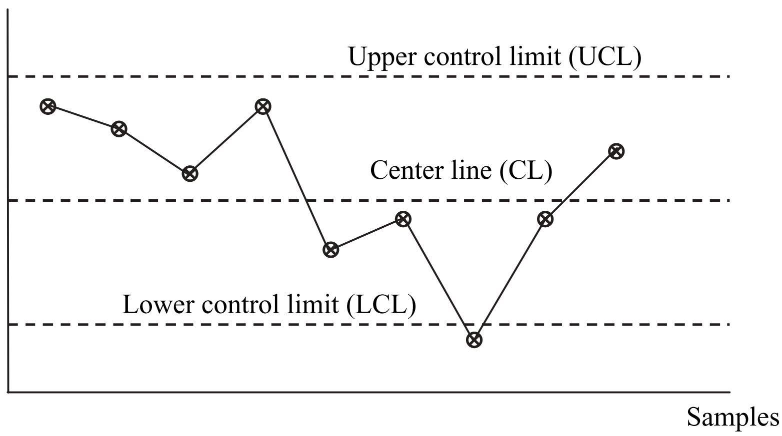

Fig. 1. A typical process control chart

Fluctuation of the points within the limits is due to variation built into the process such as design or preventative maintenance and can only be affected by changing that system. Fluctuation of the points outside of the limits comes from special causes such as people errors, unplanned outages, etc., which are not a part of the normal system or from an unlikely combination of steps. Special causes must be removed from the system to use the SPC effectively. Once this is done then the system can be described as "in control" and measurements can be taken at regular intervals to ensure that the process does not fundamentally change.

Based on the output of the process in consideration, SPC charts can be categorized into two groups. These are:

1. Variables control charts: SPC charts used to control characteristics of a product that can be measured on a continuous scale. An example of a variable would be the length or width of a product or part. Most commonly used variables control charts are X-Bar and R charts
2. Attributes control charts: SPC charts used to control which is an aspect or characteristic of a product that cannot be put on a linear scale. For example, a light bulb will either light or fail to light. "*Good/bad*" is an attribute, as is the number of defects. Examples of attributes control charts are p, np, c, and u charts.

If the process is stable, then the distribution of subgroup averages will be approximately normal. With this in mind, we can also analyze the *patterns* on the control charts to see if they might be attributed to a special cause of variation. To do this, we divide a normal distribution into zones, with each zone one standard deviation wide. Figure 2 shows the approximate percentage we expect to find in each zone from a stable process. Zone C is the area from the mean to the mean plus or minus one sigma, zone B is from plus or minus one to plus or minus two sigma, and zone A is from plus or minus two to plus or minus three sigma. Of course, any point beyond three sigma (i.e., outside of the control limit) is an indication of an out-of-control process. Since the control limits are at plus and minus three standard deviations, finding the one and two sigma lines on a control chart is as simple as dividing the distance between the grand average and either control limit into thirds, which can be done using a ruler. This divides each half of the control chart into three zones. The three zones are labeled A, B, and C as shown in Fig. 3.

Based on the expected percentages in each zone, sensitive run tests can be developed for analyzing the patterns of variation in the various zones. Reference [16] recommends four rules to identify patterns (and implicitly, points) in control charts as out-of-control. The first is the classical three-sigma rule; that is, the chart has at least one point falling outside of the three-sigma control limits. The other rules are: rule 2, two out of three consecutive points more than two sigma away from the centerline, zone A, (with the two points on the same side of the centerline); rule 3, four out of five consecutive points more than one sigma away from the centerline, zone B, (with all four on the

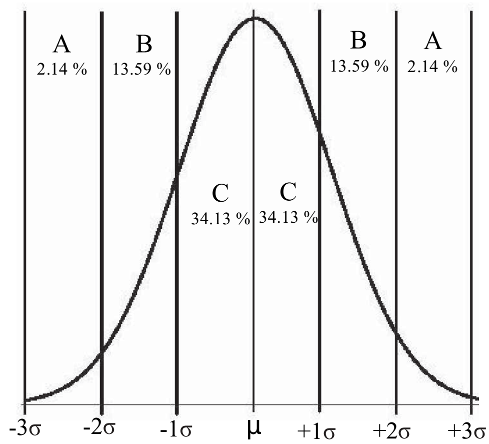

Fig. 2. Zones of normal distribution

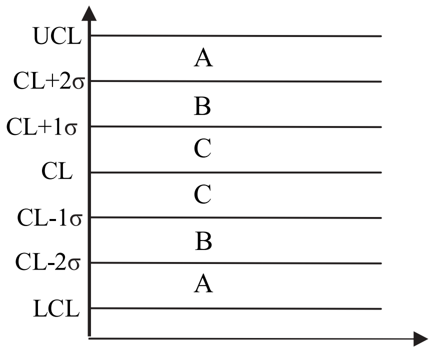

Fig. 3. Zones of a control chart

same side of the centerline); and rule 4, eight consecutive points on the same side of the centerline, zones $A+B+C$. One-sided probabilities of the rules 1, 2, 3, and 4 are calculated as 0.00135, 0.0015, 0.0027, and 0.0039, respectively.

Characteristics of interest for variables control charts can exactly be measured by the instruments and/or devices, but that for attributes cannot be measured directly and they consists of uncertainty. For this reason, fuzzy approaches to attributes control charts should be developed. In the next section, fuzzy process control charts are proposed for attributes control charts.

3 Fuzzy Process Control Charts

Even the first control chart was proposed during the 1920's by Shewhart, today they are still subject to new application areas that deserve further attention. The control charts introduced by [11] were designated to monitor processes for shifts in the mean or variance of a *single* quality characteristic. Further developments are focused on the usage of the probability and fuzzy set theories integrated with the control charts. A bibliography of control charts for attributes is presented by [17].

If the quality-related characteristics cannot be represented in numerical form, such as characteristics for appearance, softness, color, etc., then control charts for attributes are used. Product units are either classified as conforming or nonconforming, depending upon whether or not they meet specifications. The number of nonconformities (deviations from specifications) can also be counted. The binary classification into conforming and nonconforming used in the p-chart might not be appropriate in many situations where product quality does not change abruptly from satisfactory to worthless, and there might be a number of intermediate levels. Without fully utilizing such as intermediate information, the use of the p-chart usually results in poorer performance than that of the x-chart. This is evidenced by weaker detectability of process shifts and other abnormal conditions. To supplement the binary classification, several intermediate levels may be expressed in the form of linguistic terms. For example, the quality of a product can be classified by one of the following terms: "*perfect*", "*good*", "*medium*", "*poor*", or "*bad*" depending on its deviation from specifications appropriately selected continuous functions can then be used to describe the quality characteristic associated with each linguistic term.

In the literature, different procedures are proposed to monitor multinomial processes when products are classified into mutually exclusive linguistic categories. Reference [2] used fuzzy set theory as a basis for interpreting the representation of a graded degree of product conformance with quality standard. Reference [2] stressed that fuzzy economic control chart limits would be advantageous over traditional acceptance charts in that fuzzy economic control charts provide information on severity as well as the frequency of product nonconformance. References [10, 14] proposed an approach based on fuzzy set theory by assigning fuzzy sets to each linguistic term, and then combining for each sample using rules of fuzzy arithmetic and developed two approaches called *fuzzy probabilistic approach* and *membership approach*.

Apart from fuzzy probabilistic and fuzzy membership approach, [7] introduced modifications to the construction of control charts given by [13, 14]. Their study aimed at directly controlling the underlying probability distributions of the linguistic data, which were not considered by [10]. These procedures are reviewed by [18] and discussed by [9] and [1]. Reference [6] used triangular fuzzy numbers in the tests of control charts for unnatural patterns. Reference [3] proposed a neural fuzzy control chart for identifying process mean shifts. Reference [18] gave a review of statistical and fuzzy control charts based on categorical data. Reference [12] discussed different procedures of constructing control charts for linguistic data, based on fuzzy and probability theory. A comparison between fuzzy and probability approaches, based on the Average Run Length and samples under control, is conducted for real data.

Contrary to the conclusions of [10] the choice of degree of fuzziness affected the sensitivity of control charts.

Current fuzzy control charts are based on the fuzzy transformation from vague data to crisp data, and then, carried out as in the classical control charts. With the integration of the α-cut of fuzzy sets, [5] proposed α-cut fuzzy control charts.

3.1 Fuzzy *p* Control Charts

In classical p charts, products are distinctly classified as "*conformed*" or "*non-conformed*" when determining fraction rejected. In fuzzy p control charts, when categorizing products, several linguistic terms are used to denote the degree of being nonconformed product such as "*standard*", "*second choice*", "*third choice*", "*chipped*", and so on... A membership degree of being a non-conformed product is assigned to each linguistic term. Sample means for each sample group, M_j, are calculated as:

$$M_j = \frac{\sum_{i=1}^{t} k_{ij} r_i}{m_j} \tag{1}$$

where k_{ij} is the number of products categorized with the linguistic term i in the sample j, r_i is the membership degree of the linguistic term i, and m_j is the number of products in sample j. Center line, CL, is the average of the means of the n sample groups and can be determined by (2)

$$CL = \bar{M}j = \frac{\sum_{j=1}^{n} Mj}{n} \tag{2}$$

where n is the number of sample groups initially available. Since the CL is a fuzzy set, it can be represented by triangular fuzzy numbers (TFNs) whose fuzzy mode is CL, as shown in Fig. 4. Then, for each sample mean, $L_j(\alpha)$ and $R_j(\alpha)$ can be calculated using (3) and (4), respectively.

$$L_j(\alpha) = M_j\alpha \tag{3}$$

$$R_j(\alpha) = 1 - [(1 - M_j)\alpha] \tag{4}$$

Membership function of the $\bar{M}$, or CL, can be written as:

$$\mu_{M_j}(x) = \begin{cases} 0, & \text{if} \quad x \le 0 \\ \frac{x}{\bar{M}}, & \text{if} \quad 0 \le x \le \bar{M} \\ \frac{1-x}{1-\bar{M}}, & \text{if} \quad \bar{M} \le x \le 1 \\ 0, & \text{if} \quad x \ge 1 \end{cases} \tag{5}$$

Control limits for α-cut is also a fuzzy set and can be represented by TFNs. Since the membership function of CL is divided into two components, then,

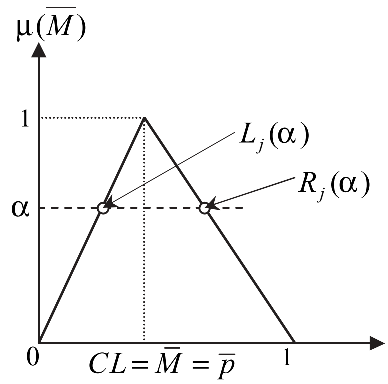

Fig. 4. TFN representation of $\bar{M}$ and M_j of the sample j

each component will have its own CL, LCL, and UCL. The membership function of the control limits depending upon the value of α is given below.

Control Limits(α)

$$= \left\{ \begin{array}{l} \left\{ \begin{array}{l} CL^L = \bar{M}\alpha \\ LCL^L = \max\left\{ CL^L - 3\sqrt{\frac{(CL^L)(1-CL^L)}{\bar{n}}},\ 0 \right\} \\ UCL^L = \min\left\{ CL^L + 3\sqrt{\frac{(CL^L)(1-CL^L)}{\bar{n}}},\ 1 \right\} \end{array} \right\},\ \text{if } 0 \le M_j \le \bar{M} \\ \\ \left\{ \begin{array}{l} CL^R = 1 - \left[\left(1 - \bar{M}\alpha\right) \alpha \right] \\ LCL^R = \max\left\{ CL^R - 3\sqrt{\frac{(CL^R)(1-CL^R)}{\bar{n}}},\ 0 \right\} \\ UCL^R = \min\left\{ CL^R + 3\sqrt{\frac{(CL^R)(1-CL^R)}{\bar{n}}},\ 1 \right\} \end{array} \right\},\ \text{if } \bar{M} \le M_j \le 1 \end{array} \right. \tag{6}$$

where $\bar{n}$ is the average sample size (ASS). When the ASS is used, the control limits do not change with the sample size. Hence, the control limits for all samples are the same. A general illustration of these control limits is shown in Fig. 5.

For the variable sample size (VSS), $\bar{n}$ should be replaced by the size of the jth sample n_j. Hence, control limits change for each sample depending upon the size of the sample. Therefore, each sample has its own control limits. The decision that whether process is *in control* (1) or *out of control* (0) for both ASS and VSS is as follows:

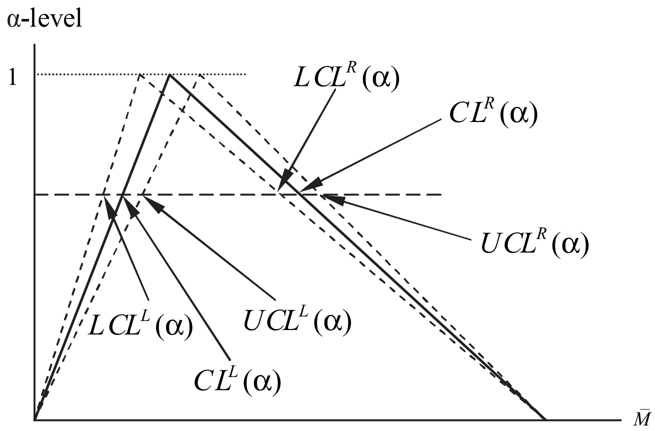

Fig. 5. Illustration of the α-cut control limits (ASS)

$$\text{Process Control} = \begin{cases} 1, & \text{if } LCL^L(\alpha) \le L_j(\alpha) \le UCL^L(\alpha) \wedge LCL^R(\alpha) \\ & \quad \le R_j(\alpha) \le UCL^R(\alpha) \\ 0, & \text{otherwise} . \end{cases} \tag{7}$$

The value of α-cut is decided with respect to the tightness of inspection such that for a tight inspection, α values close to 1 may be used. As can be seen from Fig. 5, while α reduces to 0 (decreasing the tightness of inspection), the range where the process is *in control* (difference between UCL and LCL) increases.

3.2 Fuzzy *c* Control Charts

In the crisp case, control limits for number of nonconformities are calculated by the (8–10).

$$CL = \bar{c} \tag{8}$$
$$LCL = \bar{c} - 3\sqrt{\bar{c}} \tag{9}$$
$$UCL = \bar{c} + 3\sqrt{\bar{c}} \tag{10}$$

where $\bar{c}$ is the mean of the nonconformities. In the fuzzy case, each sample, or subgroup, is represented by a trapezoidal fuzzy number (a, b, c, d) or a triangular fuzzy number (a, b, d) as shown in Fig. 6. Note that a trapezoidal fuzzy number becomes triangular when $b = c$. For the ease of representation and calculation, a triangular fuzzy number is also represented as trapezoidal by (a, b, b, d) or (a, c, c, d). Center line, $\widetilde{CL}$, given in (8), is the mean of fuzzy samples, and it is shown as ($\bar{a}$, $\bar{b}$, $\bar{c}$, $\bar{d}$) where $\bar{a}$, $\bar{b}$, $\bar{c}$, and $\bar{d}$ are the arithmetic means of the a, b, c, and d, respectively:

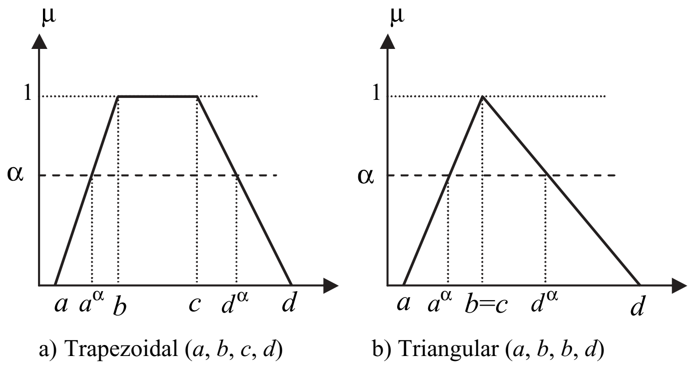

Fig. 6. Representation of a sample by trapezoidal and/or triangular fuzzy numbers

$$\widetilde{CL} = \left(\frac{\sum_{j=1}^{n} a_j}{n}, \frac{\sum_{j=1}^{n} b_j}{n}, \frac{\sum_{j=1}^{n} c_j}{n}, \frac{\sum_{j=1}^{n} d_j}{n} \right) = \left(\bar{a}, \bar{b}, \bar{c}, \bar{d}\right) \tag{11}$$

where n is the number of fuzzy samples.

Since the $\widetilde{CL}$ is a fuzzy set, it can be represented by a fuzzy number whose fuzzy mode (multimodal) is the closed interval of $[\bar{b}, \bar{c}]$. $\widetilde{CL}$, $\widetilde{LCL}$, and $\widetilde{UCL}$ are calculated using (12–14).

$$\widetilde{CL} = \left(\bar{a}, \bar{b}, \bar{c}, \bar{d}\right) = (CL_1, CL_2, CL_3, CL_4) \tag{12}$$

$$\begin{aligned} \widetilde{LCL}^{\alpha} &= \widetilde{CL}^{\alpha} - 3\sqrt{\widetilde{CL}^{\alpha}} = \left(\bar{a}, \bar{b}, \bar{c}, \bar{d}\right) - 3\sqrt{\left(\bar{a}, \bar{b}, \bar{c}, \bar{d}\right)} \\ &= \left(\bar{a} - 3\sqrt{\bar{d}}, \bar{b} - 3\sqrt{\bar{c}}, \bar{c} - 3\sqrt{\bar{b}}, \bar{d} - 3\sqrt{\bar{a}}\right) \\ &= (LCL_1, LCL_2, LCL_3, LCL_4) \end{aligned} \tag{13}$$

$$\begin{aligned} \widetilde{UCL} &= \widetilde{CL} + 3\sqrt{\widetilde{CL}} = (\bar{a}, \bar{b}, \bar{c}, \bar{d}) + 3\sqrt{(\bar{a}, \bar{b}, \bar{c}, \bar{d})} \\ &= \left(\bar{a} + 3\sqrt{\bar{a}}, \bar{b} + 3\sqrt{\bar{b}}, \bar{c} + 3\sqrt{\bar{c}}, \bar{d} + 3\sqrt{\bar{d}}\right) \\ &= (UCL_1, UCL_2, UCL_3, UCL_4) \end{aligned} \tag{14}$$

An α-cut is a nonfuzzy set which comprises all elements whose membership is greater than or equal to α. Applying α-cuts of fuzzy sets (Fig. 4) values of a^{α} and d^{α} are determined by (15) and (16), respectively.

$$a^{\alpha} = a + \alpha(b - a) \tag{15}$$

$$d^{\alpha} = d - \alpha(d - c) \tag{16}$$

Using α-cut representations, fuzzy control limits can be rewritten as given in (17–19).

$$\widetilde{CL}^{\alpha} = (\bar{a}^{\alpha}, \bar{b}, \bar{c}, \bar{d}^{\alpha}) = (CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha}) \tag{17}$$

$$\begin{aligned} \widetilde{LCL}^{\alpha} &= \widetilde{CL}^{\alpha} - 3\sqrt{\widetilde{CL}^{\alpha}} = (\bar{a}^{\alpha}, \bar{b}, \bar{c}, \bar{d}^{\alpha}) - 3\sqrt{(\bar{a}^{\alpha}, \bar{b}, \bar{c}, \bar{d}^{\alpha})} \\ &= (\bar{a}^{\alpha} - 3\sqrt{\bar{d}^{\alpha}}, \bar{b} - 3\sqrt{\bar{c}}, \bar{c} - 3\sqrt{\bar{b}}, \bar{d}^{\alpha} - 3\sqrt{\bar{a}^{\alpha}}) \\ &= (LCL_1^{\alpha}, LCL_2, LCL_3, LCL_4^{\alpha}) \end{aligned} \tag{18}$$

$$\begin{aligned} \widetilde{UCL}^{\alpha} &= \widetilde{CL}^{\alpha} + 3\sqrt{\widetilde{CL}^{\alpha}} = (\bar{a}^{\alpha}, \bar{b}, \bar{c}, \bar{d}^{\alpha}) + 3\sqrt{(\bar{a}^{\alpha}, \bar{b}, \bar{c}, \bar{d}^{\alpha})} \\ &= \left(\bar{a}^{\alpha} + 3\sqrt{\bar{a}^{\alpha}}, \bar{b} + 3\sqrt{\bar{b}}, \bar{c} + 3\sqrt{\bar{c}}, \bar{d}^{\alpha} + 3\sqrt{\bar{d}^{\alpha}}\right) \\ &= (UCL_1^{\alpha}, UCL_2, UCL_3, UCL_4^{\alpha}) \end{aligned} \tag{19}$$

Results of these equations can be illustrated as in Fig. 7. To retain the standard format of control charts and to facilitate the plotting of observations on the chart, it is necessary to convert the fuzzy sets associated with linguistic

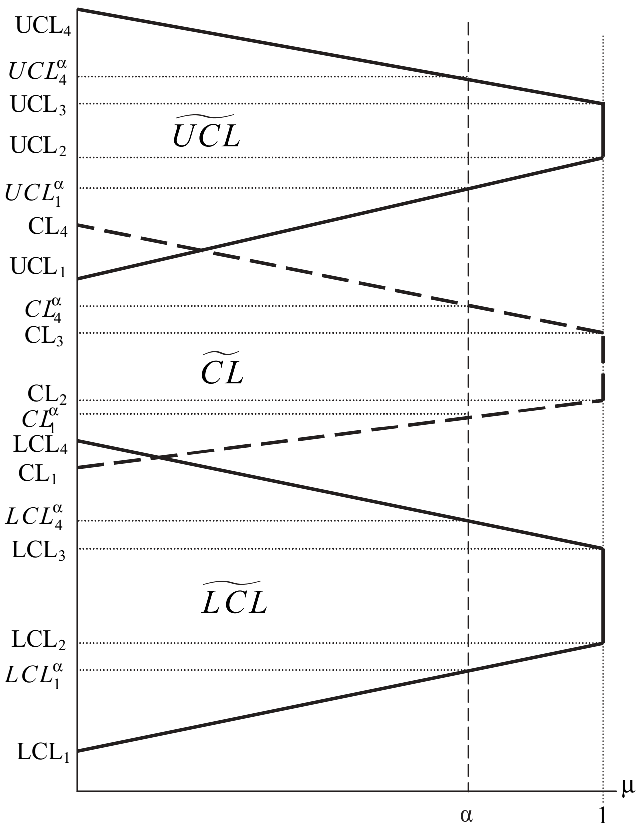

Fig. 7. Representation of fuzzy control limits

values into scalars referred to as *representative values*. This conversion may be performed in a number of ways as long as the result is intuitively representative of the range of the base variable included in the fuzzy set. Four ways, which are similar in principle to the measures of central tendency used in descriptive statistics, are *fuzzy mode*, *α-level fuzzy midrange*, *fuzzy median*, and *fuzzy average*. It should be pointed out that there is no theoretical basis supporting any one specifically and the selection between them should be mainly based on the ease of computation or preference of the user [14]. Conversion of fuzzy sets into crisp values results in loss of information in linguistic data. To retain information of linguistic data we prefer to keep fuzzy sets as themselves and to compare fuzzy samples with the fuzzy control limits. For this reason, a direct fuzzy approach (DFA) based on the area measurement is proposed for the fuzzy control charts. α-level fuzzy control limits, $\widetilde{UCL}^{\alpha}$, $\widetilde{CL}^{\alpha}$, and, $\widetilde{LCL}^{\alpha}$, can be determined by fuzzy arithmetic as shown in (20–22).

$$\widetilde{CL}^{\alpha} = (CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha}) \tag{20}$$

$$\begin{aligned}\widetilde{LCL}^{\alpha} &= \widetilde{CL}^{\alpha} - 3\sqrt{\widetilde{CL}^{\alpha}} = (CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha}) \\ &\quad - 3\sqrt{(CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha})} \\ &= \left(CL_1^{\alpha} - 3\sqrt{CL_4^{\alpha}}, CL_2 - 3\sqrt{CL_3}, CL_3 - 3\sqrt{CL_2}, CL_4^{\alpha} - 3\sqrt{CL_1^{\alpha}}\right) \\ &= (LCL_1^{\alpha}, LCL_2, LCL_3, LCL_4^{\alpha})\end{aligned} \tag{21}$$

$$\begin{aligned}\widetilde{UCL}^{\alpha} &= \widetilde{CL}^{\alpha} + 3\sqrt{\widetilde{CL}^{\alpha}} = (CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha}) \\ &\quad + 3\sqrt{(CL_1^{\alpha}, CL_2, CL_3, CL_4^{\alpha})} \\ &= \left(CL_1^{\alpha} + 3\sqrt{CL_1^{\alpha}}, CL_2 + 3\sqrt{CL_2}, CL_3 + 3\sqrt{CL_3}, CL_4^{\alpha} + 3\sqrt{CL_4^{\alpha}}\right) \\ &= (UCL_1^{\alpha}, UCL_2, UCL_3, UCL_4^{\alpha})\end{aligned} \tag{22}$$

where,

$$CL_1^{\alpha} = CL_1 + \alpha\,(CL_2 - CL_1) \tag{23}$$

$$CL_4^{\alpha} = CL_4 - \alpha\,(CL_4 - CL_3) \tag{24}$$

Decision about whether the process is in control can be made according to the percentage area of the sample which remains inside the $\widetilde{UCL}$ and/or $\widetilde{LCL}$ defined as fuzzy sets. When the fuzzy sample is completely involved by the fuzzy control limits, the process is said to be "*in-control*". If a fuzzy sample is totally excluded by the fuzzy control limits, the process is said to be "*out of control*". Otherwise, a sample is partially included by the fuzzy control limits. In this case, if the percentage area which remains inside the fuzzy control limits (β_j) is equal or greater than a predefined acceptable percentage (β), then the process can be accepted as "*rather in-control*"; otherwise it can be stated as "*rather out of control*". Possible decisions resulting from DFA are illustrated in Fig. 8. Parameters for determination of the sample area outside the control

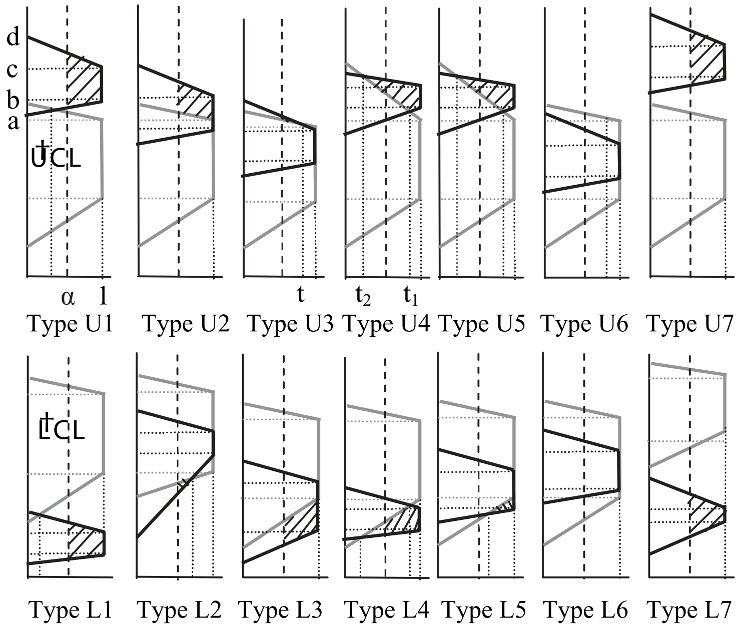

Fig. 8. Illustration of all possible sample areas outside the fuzzv control limits at α-level cut

limits for α-level fuzzy cut are LCL_1, $LCL_2, UCL_3, UCL_4, a, b, c, d$, and α. The shape of the control limits and fuzzy sample are formed by the lines of $\overline{LCL_1 LCL_2}$, and $\overline{UCL_1 UCL_2}, \overline{ab}$, $\overline{cd}$. A flowchart to calculate area of the fuzzy sample outside the control limits is given in Fig. 9. Sample area above the upper control limits, A^U_{out}, and sample area falling below the lower control limits, A^L_{out}, are calculated. Equations to compute A^U_{out} and A^L_{out} are given in Appendix A. Then, total sample area outside the fuzzy control limits, A_{out}, is the sum of the areas below fuzzy lower control limit and above fuzzy upper control limit. Percentage sample area within the control limits is calculated as given in (25).

$$\beta_j^\alpha = \frac{S_j^\alpha - A^\alpha_{\text{out},j}}{S_j^\alpha} \tag{25}$$

where S_j^α is the sample area at α-level cut.

DFA provides the possibility of obtaining linguistic decisions like "rather in control" or "rather out of control". Further intermediate levels of process control decisions are also possible by defining in stages. For instance, it may be defined as given below which is more distinguished.

$$\text{Process Control} = \begin{cases} \text{in control,} & 0.85 \le \beta_j \le 1 \\ \text{rather in control,} & 0.60 \le \beta_j < 0.85 \\ \text{rather out of control,} & 0.10 \le \beta_j < 0.60 \\ \text{out of control,} & 0 \le \beta_j < 0.10 \end{cases} \tag{26}$$

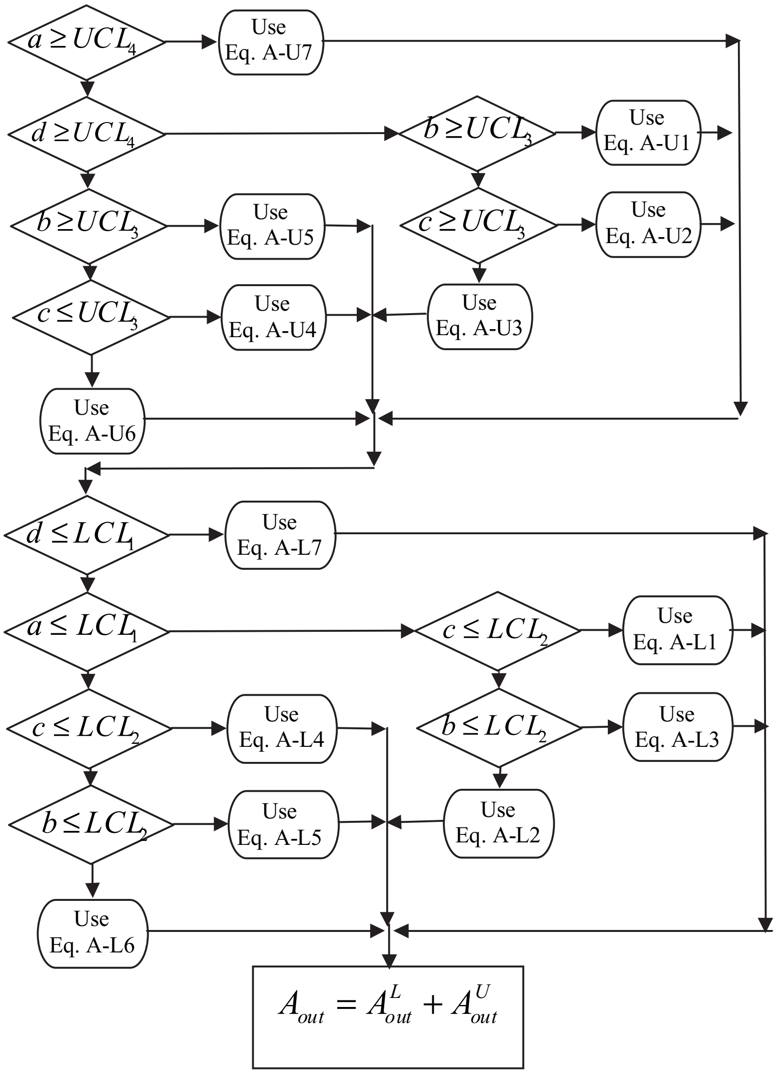

Fig. 9. Flowchart to cumpute the area outside the fuzzy control limits

4 Fuzzy Process Control with Intelligent Data Mining

Analysis of fuzzy unnatural patterns for fuzzy control charts is necessary to develop. Run rules are based on the premise that a specific run of data has a low probability of occurrence in a completely random stream of data. If a run occurs, then this must mean that something has changed in the process to produce a nonrandom or unnatural pattern. Based on the expected percentages in each zone, sensitive run tests can be developed for analyzing the patterns of variation in the various zones. The formula for calculating the probability of a fuzzy event A is a generalization of the probability theory:

$$P(A) = \begin{cases} \int \mu_A(x) P_x(x) dx, & \text{if } X \text{ is continuous} \\ \sum_i \int \mu_A(x_i) P_x(x_i), & \text{if } X \text{ is discrete} \end{cases} \tag{27}$$

where P_X denotes the probability distribution function of X. The membership degree of a sample to belong to a region is directly related to its percentage area falling in that region, and therefore, it is continuous. For example, a fuzzy sample may be in zone B with a membership degree of 0.4 and in zone C with a membership degree of 0.6. While counting points in zone B, that point is counted as 0.4.

Based on the Western Electric rules, the following fuzzy unnatural pattern rules can be defined.

Rule 1: Any fuzzy data falling outside the three-sigma control limits with a ratio (25) of more than predefined percentage (β) of sample area at desired α-level:

$$\mu_1 = \begin{cases} 0, & 0.85 \le x \le 1 \\ (x - 0.60)/0.25, & 0.60 \le x \le 0.85 \\ (x - 0.10)/0.50, & 0.10 \le x \le 0.60 \\ 1, & 0 \le x \le 0.10 \end{cases} \tag{28}$$

where x is the ratio of fuzzy data falling outside the three-sigma control limits.

Rule 2: A total membership degree *around* 2 from 3 consecutive points in zone A or beyond. Probability of a sample being in zone A (0.0214) or beyond (0.00135) is 0.02275. Let membership function for this rule be defined as follows:

$$\mu_2 = \begin{cases} 0, & 0 \le x \le 0.59 \\ (x - 0.59)/1.41, & 0.59 \le x \le 2 \\ 1, & 2 \le x \le 3 \end{cases} \tag{29}$$

Using the membership function above, fuzzy probability given in (27) can be rewritten as follows:

$$\begin{aligned} \int_0^3 \mu_2(x) P_2(x) &= \int_0^{x_1} \mu_2(x) P_2(x) + \int_{x_1}^{x_2} \mu_2(x) P_2(x) + \int_{x_2}^{3} \mu_2(x) P_2(x) \\ &= \int_{x_1}^{x_2} \mu_2(x) P_2(x) + \int_{x_2}^{3} \mu_2(x) P_2(x) \end{aligned} \tag{30}$$

where,

$$P_x(x) = P_x\left(z \ge \frac{x - np}{\sqrt{npq}}\right) \tag{31}$$

To integrate the (30), membership function is divided into sections, each with a 0.05 width, and $\mu_2(x)P_x(x)$ values for each section are summed. For $x_1 = 0.59$ and $x_2 = 2$, the probability of the fuzzy event, rule 2, is determined as 0.0015, which corresponds to the crisp case of this rule. In the following rules, the membership functions are set in the same way.

Rule 3: A total membership degree *around* 4 from 5 consecutive points in zone C or beyond with the membership function (degree of unnaturalness) given below:

$$\mu_3 = \begin{cases} 0, & 0 \le x \le 2.42 \\ (x-2.42)/1.58, & 2.42 \le x \le 4 \\ 1, & 4 \le x \le 5 \end{cases} \tag{32}$$

The fuzzy probability for this rule is calculated as 0.0027.

Rule 4: A total membership degree *around* 8 from 8 consecutive points on the same side of the centerline:

$$\mu_4 = \begin{cases} 0, & 0 \le x \le 2.54 \\ (x-2.54)/5.46, & 2.54 \le x \le 8 \end{cases} \tag{33}$$

The fuzzy probability for the rule above is then determined as 0.0039. Probability of each fuzzy rule (event) above depends on the definition of the membership function which is subjectively defined with respect to the classical probabilities for unnatural patterns.

5 A Numerical Example for Fuzzy *c* Control Charts

Samples of 200 units are taken every 4 hours to control number of nonconformities. Data collected from 30 subgroups shown in Table 2 are linguistic such as *"approximately 30"* or *"between 25 and 30"*.

The linguistic expressions in Table 1 are represented by fuzzy numbers as shown in Table 3. These numbers are subjectively identified by the quality control expert who also sets $\alpha = 0.60$ and minimum acceptable ratio as $\beta = 0.70$. Quality control expert also set the acceptable membership degree of unnaturalness as 0.95, that is, when a sample refers to an unnatural sample with respect to any rule, it should refer a membership degree of unnaturalness more than 0.95 with respect to the membership function defined for that rule.

Using (5–7), $\widetilde{CL}$, $\widetilde{LCL}$, and $\widetilde{UCL}$ are determined as follows:

$$\widetilde{CL} = (18.13, 22.67, 26.93, 32.07)$$
$$\widetilde{LCL} = (1.15, 7.10, 12.65, 19.29)$$
$$\widetilde{UCL} = (30.91, 36.95, 42.50, 49.05)$$

Table 1. Number of nonconformities for 30 subgroups

No	Approximately	Between	No	Approximately	Between
1	30		16	40	
2		20–30	17		32–50
3		5–12	18	39	
4	6		19		15–21
5	38		20	28	
6		20–24	21		32–35
7		4–8	22		10–25
8		36–44	23	30	
9		11–15	24	25	
10		10–13	25		31–41
11	6		26		10–25
12	32		27		5–14
13	13		28		28-35
14		50–52	29		20–25
15		38–41	30	8	

Table 2. Fuzzy number (a, b, c, d) representation of 30 subgroups

No	a	b	c	d	No	a	b	c	d
1	25	30	30	35	16	33	40	40	44
2	15	20	30	35	17	28	32	50	60
3	4	5	12	15	18	33	39	39	43
4	3	6	6	8	19	12	15	21	38
5	32	38	38	45	20	23	28	28	36
6	16	20	24	28	21	28	32	35	42
7	3	4	8	12	22	14	18	28	33
8	27	36	44	50	23	24	30	30	34
9	9	11	15	20	24	20	25	25	31
10	7	10	13	15	25	25	31	41	46
11	3	6	6	10	26	7	10	25	28
12	27	32	32	37	27	3	5	14	20
13	11	13	13	15	28	23	28	35	38
14	39	50	52	55	29	17	20	25	29
15	28	38	41	45	30	5	8	8	15
		Average				18.13	22.67	26.93	32.07

Table 3. Fuzzy zones calculated for the example problem

Zone	a	b	c	d
UCL^{α}	34.53	36.95	42.50	45.12
$+2\ \sigma$	29.97	32.19	37.31	39.74
$+1\ \sigma$	25.41	27.43	32.12	34.37
CL^{α}	20.85	22.67	26.93	28.99
$-1\ \sigma$	15.47	17.48	22.17	24.43
$-2\ \sigma$	10.10	12.29	17.41	19.87
LCL^{α}	4.72	7.10	12.65	15.31

Applying α-cut of 0.60, values of $\widetilde{CL}^{\alpha=0.60}$, $\widetilde{LCL}^{\alpha=0.60}$, and $\widetilde{UCL}^{\alpha=0.60}$ are calculated. (10–12)

$$\widetilde{CL}^{\alpha=0.60} = (20.85, 22.67, 26.93, 28.99)$$
$$\widetilde{LCL}^{\alpha=0.60} = (4.72, 7.10, 12.65, 15.31)$$
$$\widetilde{UCL}^{\alpha=0.60} = (34.53, 36.95, 42.50, 45.12)$$

Membership functions in (28, 29, 32), and (33) are used for the rules 1–4 as μ_1, μ_2, μ_3, and μ_4, respectively. These membership functions set the degree of unnaturalness for each rule. As an example, when a total membership degree of 1.90 is calculated for the rule 2, its degree of unnaturalness is determined from μ_2 as 0.9291. Probabilities of these fuzzy events are calculated using normal approach to binomial distribution.

In order to make calculations easy and mine our sample database for unnaturalness a computer program is coded using Fortran 90 programming language. Table 4 gives total membership degrees of the fuzzy samples and those of unnaturalness in zones.

As can be seen from Table 4, sample 14 is beyond the $\pm 3\sigma$ limits and shows an out of control situation. Considering samples 14, 15, and 16 for the rule 2, total membership degree is calculated as 2, which refers an unnatural pattern with unnaturalness degree of 1. Then, only the last sample, sample 16, is marked as unnatural pattern and necessary actions should be taken at this stage. The mining is the restarted with sample 17. With the sample 18, total membership degree for rule 2 reaches to 1.61 and degree of unnaturalness determined from μ_2 is 0.72. Since minimum acceptable membership degree of unnaturalness for this problem is set to 0.95, it is not treated as an unnatural pattern. Sample 10 denotes unnaturalness with respect to the rule 2 applied to the lower side of the control chart. There is no sample indicating unnaturalness with a degree more than 0.95 according to the rules 3 and 4.

Table 4. Total membership degrees of the fuzzy samples and degree of unnaturalness in zones

Sample No	Beyond $\pm 3\sigma$	In or Above Fuzzy CL			In or Below Fuzzy CL		
		Rule 2	Rule 3	Rule 4	Rule 2	Rule 3	Rule 4
1	0.00	0.24	1	1	0	0	0.06
2	0.00	0.04	0.38	0.77	0.03	0.36	0.75
3	0.14	0	0	0	0.86	0.86	0.86
4	0.32	0	0	0	0.68	0.68	0.68
5	0.00	1	1	1	0	0	0
6	0.00	0	0	0.54	0.05	0.73	1
7	0.42	0	0	0	0.58	0.58	0.58
8	0.13	0.87	0.87	0.87	0	0	0
9	0.00	0	0	0	1	1	1
10	0.00	0	0	0	**1 ($\mu^* = 1$)**	1	1
11	0.26	0	0	0	0.74	0.74	0.74
12	0.00	0.96	1	1	0	0	0
13	0.00	0	0	0	1	1	1
14	**1.00**	0	0	0	0	0	0
15	0.00	1	1	1	0	0	0
16	0.00	**1 ($\mu^*=1$)**	1	1	0	0	0
17	0.39	0.61	0.61	0.61	0	0	0
18	0.00	1 ($\mu^* = 0.72$)	1	1	0	0	0
19	0.00	0	0.03	0.28	0.42	0.87	1
20	0.00	0.05	1	1	0	0	0.42
21	0.00	0.99	1	1	0	0	0
22	0.00	0	0.22	0.61	0.13	0.52	0.91
23	0.00	0.17	1	1	0	0	0.11
24	0.00	0	0.2	1	0	0.11	1
25	0.00	0.9	1	1	0	0	0
26	0.00	0	0.01	0.24	0.57	0.86	1
27	0.12	0	0	0	0.88	0.88	0.88
28	0.00	0.53	1	1	0	0	0.13
29	0.00	0	0.03	0.63	0.02	0.61	1
30	0.00	0	0	0	1	1	1

* unnatural sample with the corresponding degree of unnaturalness defined by the membersip functions for each rule.

6 Conclusions

In this chapter, fuzzy process control charts for attributes have been developed. Well-known Western Electric rules for examining unnaturalness are fuzzified using probability of fuzzy events and searched with data mining. A linguistic data of 30 samples have been used for illustration purposes. For larger data sets, unnaturalness can be mined using the same procedure. There are other unnatural pattern rules defined in the literature. These rules can also be examined under fuzziness. Some new rules can be added to the existing rules. When rules for unnaturalness are defined for longer runs, the usage of an intelligent data mining procedure is inevitable.

Appendix

Equations to compute sample area outside the control the limits.

$$A_{\text{out}}^{U} = \frac{1}{2}\left[\left(d^{\alpha} - UCL_4^{\alpha}\right) + \left(d^{t} - UCL_4^{t}\right)\right](\max(t-\alpha, 0)) + \frac{1}{2}\left[(d^{z} - a^{z}) + (c-b)\right](\min(1-t, 1-\alpha)) \quad \text{(A-U1)}$$

where,

$$t = \frac{UCL_4 - a}{(b-a)+(c-b)} \text{ and } z = \max(t, \alpha)$$

$$A_{\text{out}}^{U} = \frac{1}{2}\left[\left(d^{\alpha} - UCL_4^{\alpha}\right) + (c - UCL_3)\right](1-\alpha) \quad \text{(A-U2)}$$

$$A_{\text{out}}^{U} = \frac{1}{2}(d^{\alpha} - UCL_4^{\alpha})(\max(t-\alpha, 0)) \quad \text{(A-U3)}$$

$$A_{\text{out}}^{U} = \frac{1}{2}\left[(c - UCL_3) + \left(d^{z} - UCL_4^{z}\right)\right](\max(1-t, 1-\alpha)) \quad \text{(A-U4)}$$

where

$$t = \frac{UCL_4 - d}{(UCL_4 - UCL_3) - (d-c)} \text{ and } z = \max(t, \alpha)$$

$$A_{\text{out}}^{U} = \frac{1}{2}\left[\left(d^{z_2} - UCL_4^{z_2}\right) + \left(d^{t_1} - UCL_4^{t_1}\right)\right] \times (\min(\max(t_1-\alpha, 0), t_1 - t_2)) + \frac{1}{2}\left[\left(d^{z_1} - a^{z_1}\right) + (c-b)\right](\min(1-t_1, 1-\alpha)) \quad \text{(A-U5)}$$

where

$$t_1 = \frac{UCL_4 - a}{(b-a) + (UCL_4 - UCL_3)},$$

$$t_2 = \frac{UCL_4 - d}{(UCL_4 - UCL_3) - (d-c)}$$

$$z_1 = \max(\alpha, t_1), \text{and } z_2 = \max(\alpha, t_2)$$

$$A_{\text{out}}^{U} = 0 \quad \text{(A-U6)}$$

$$A_{\text{out}}^{U} = \frac{1}{2}[(d^{\alpha} - a^{\alpha}) + (c-b)](1-\alpha) \quad \text{(A-U7)}$$

$$A_{\text{out}}^L = \frac{1}{2}[(LCL_1^\alpha - a^\alpha) + (LCL_1^t - a^t)](\max(t-\alpha,0)) + \frac{1}{2}[(d^z - a^z) + (c-b)](\min(1-t,1-\alpha)) \quad \text{(A-L1)}$$

where

$$t = \frac{d - LCL_1}{(LCL_2 - LCL_1) + (d-c)} \text{ and } z = \max(\alpha, t)$$

$$A_{\text{out}}^L = \frac{1}{2}\left[\left(d^\alpha - a^\alpha\right) + (c-b)\right](1-\alpha) \quad \text{(A-L2)}$$

$$A_{\text{out}}^L = \frac{1}{2}(LCL_1^\alpha - a^\alpha) + (LCL_2 - b)](1-\alpha) \quad \text{(A-L3)}$$

$$A_{\text{out}}^L = \frac{1}{2}\left[\left(LCL_1^{z_2} - a^{z_2}\right) + \left(LCL_1^{t_1} - a^{t_1}\right)\right] \times (\min(\max(t_1-\alpha,0), t_1 - t_2)) + \frac{1}{2}\left[\left(d^{z_1} - a^{z_1}\right) + (c-b)\right](\min(1-t,1-\alpha)) \quad \text{(A-L4)}$$

where

$$t_1 = \frac{d - LCL_1}{(LCL_2 - LCL_1) + (d-c)}, t_2 = \frac{a - LCL_1}{(LCL_2 - LCL_1) + (b-a)}$$

$$z_1 = \max(\alpha, t_1), \text{and } z_2 = \max(\alpha, t_2)$$

$$A_{\text{out}}^L = \frac{1}{2}\left[\left(LCL_4^z - a^z\right) + (LCL_2 - b)\right](\min(1-t,1-\alpha)) \quad \text{(A-L5)}$$

where

$$t = \frac{a - LCL_1}{(LCL_2 - LCL_1) - (b-a)}, \text{and } z = \max(\alpha, t)$$

$$A_{\text{out}}^L = 0 \quad \text{(A-L6)}$$

$$A_{\text{out}}^L = \frac{1}{2}[(d^\alpha - a^\alpha) + (c-b)](1-\alpha) \quad \text{(A-L7)}$$

References

1. Asai, K. (1995), Fuzzy Systems for Management, IOS Press. 320
2. Bradshaw, C.W. (1983), A fuzzy set theoretic interpretation of economic control limits, *European Journal of Operational Research*, 13(4), 403–408. 320
3. Chang S.I., Aw, C.A. (1996), A neural fuzzy control chart for detecting and classifying process mean shifts, *International Journal of Production Research*, 34 (8), 2265–2278. 320
4. Fayyad, U. (1998), Mining Databases: Towards Algorithms for Knowledge Discovery, *IEEE Computer Society*, Special Issue on Mining Databases, Vol. 21, No. 1, pp. 39–48. 316

5. Gülbay, M., Kahraman, C., Ruan, D. (2004), α-cut fuzzy control charts for linguistic data, *International Journal of Intelligent Systems*, 19, 12, pp. 1173–1195. 321
6. Kahraman, C., Tolga, E., Ulukan, Z. (1995), Using triangular fuzzy numbers in the tests of control charts for unnatural patterns, in Proceedings of *INRIA / IEEE Conference on Emerging Technologies and Factory Automation*, October, 10-13, Paris-France, Vol. 3, pp. 291–298. 320
7. Kanagawa, A., Tamaki, F., Ohta, H. (1993), Control charts for process average and variability based on linguistic data, *International Journal of Production Research*, 2, 913–922. 320
8. Klemettinen, M., Mannila, H., Toivonen, H. (1997), A datamining methodology and its application to semi-automatic knowledge acquisition, *In DEXA'97 Workshop*, September 1–5, 1997. IEEE Computer Society Press. 316
9. Laviolette, M., Seaman, J.W., Barrett, J.D., Woodall, W.H. (1995), A probabilistic and statistical view of fuzzy methods, with discussion., *Technometrics*, 37, pp. 249–292. 320
10. Raz, T., Wang, J.-H. (1990), Probabilistic and membership approaches in the construction of control charts for linguistic data, *Production Planning and Control*, 1, 147–157. 320, 321
11. Shewhart, W.A. (1931), Economic control of quality of manufactured product, *D. Van Nostrand,In.c,* Princeton, N.J. 319
12. Taleb, H., Limam, M. (2002), On fuzzy and probabilistic control charts, *International Journal of Production Research*, 40 (12), 2849–2863. 320
13. Wang, J.-H., Raz, T. (1988), Applying fuzzy set theory in the development of quality control charts, *International Industrial Engineering Conference proceedings,* Orlando, FL, 30–35. 320
14. Wang, J.-H., Raz, T. (1990), On the construction of control charts using linguistic variables, *International Journal of Production Research*, 28, 477–487. 320, 326
15. Wang, J.-H., Chen, C.-H. (1995), Economic statistical *np*-control chart designs based on fuzzy optimization, *International Journal of Quality and Reliability Management*, 12(1), 88–92.
16. Western Electric Company, Bonnie B. Small, Chairman of the Writing Committee (1956), *StatisticalQuality Control Handbook*, Indianapolis, IN: AT&T Technologies (Select Code 700–444, P.O. Box 19901, Indianapolis 46219). 318
17. Woodall, W. (1997), Control charts based on attribute data: bibliography and review., *Journal of Quality Technology.,* 29, pp. 172–183. 320
18. Woodall, W., Tsui, K.-L., Tucker, G.L. (1997), A review of statistical and fuzzy control charts based on categorical data, *Frontiers in Statistical Quality Control 5 (Heidelberg, Germany: Physica-Verlag)*, pp. 83–89. 320