Available online at www.sciencedirect.com
ScienceDirect
Computers & Industrial Engineering 51 (2006) 726–742
computers & industrial engineering
www.elsevier.com/locate/dsw

# Feature-based recognition of control chart patterns ☆

Susanta Kumar Gauri [a], Shankar Chakraborty [b,*]

[a] SQC and OR Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, India
[b] Department of Production Engineering, Jadavpur University, Kolkata 700032, India

Received 19 July 2005; received in revised form 8 May 2006; accepted 17 July 2006
Available online 28 September 2006

**Abstract**

Control charts primarily in the form of $\overline{X}$ chart are widely used to identify the situations when control actions will be needed for manufacturing systems. Various types of patterns are observed in control charts. Identification of these control chart patterns (CCPs) can provide clues to potential quality problems in the manufacturing process. Each type of control chart pattern has its own geometric shape and various related features can represent this shape. Feature-based approaches can facilitate efficient pattern recognition since extracted shape features represent the main characteristics of the patterns in a condensed form. In this paper, a set of eight new features, extraction of which does not call for utilizing the experience and skill of the user in any form, is presented. Two feature-based approaches using heuristics and artificial neural network (ANN) are developed, which are capable of recognizing eight most commonly observed CCPs including stratification and systematic patterns. Relative performances of the feature-based heuristic and feature-based ANN recognizers are extensively studied using synthetic pattern data. The feature-based ANN recognizer results in better recognition performance and generalization compared to the feature-based heuristic recognizer.


*Keywords:* Control chart patterns; Features; Pattern recognition; Feature-based heuristic recognizer; Feature-based ANN recognizer; Classification performance

## 1. Introduction

The primary objective of process quality control is to achieve and maintain an acceptable level of the desired process quality characteristic consistently. In this connection, accurate monitoring and control of the manufacturing system is very important. Control charts predominantly in the form of $\overline{X}$ chart are widely used to identify situations when control actions will be needed for the manufacturing systems. Commonly eight types of control chart patterns (CCPs), as shown in Fig. 1, are encountered in different manufacturing environments. The patterns can be classified as natural/normal and unnatural/abnormal (Grant & Leavenworth, 1996; Montgomery, 2000). The basic significance of a natural pattern is that it indicates a process under

☆ Processed by area editor E.A. Elsayed

* Corresponding author. Tel.: +91 33 2414 6153; fax: +91 33 2414 6414/2413 7121.
*E-mail address:* s_chakraborty00@yahoo.co.in (S. Chakraborty).

0360-8352/$ - see front matter 
doi:10.1016/j.cie.2006.07.013

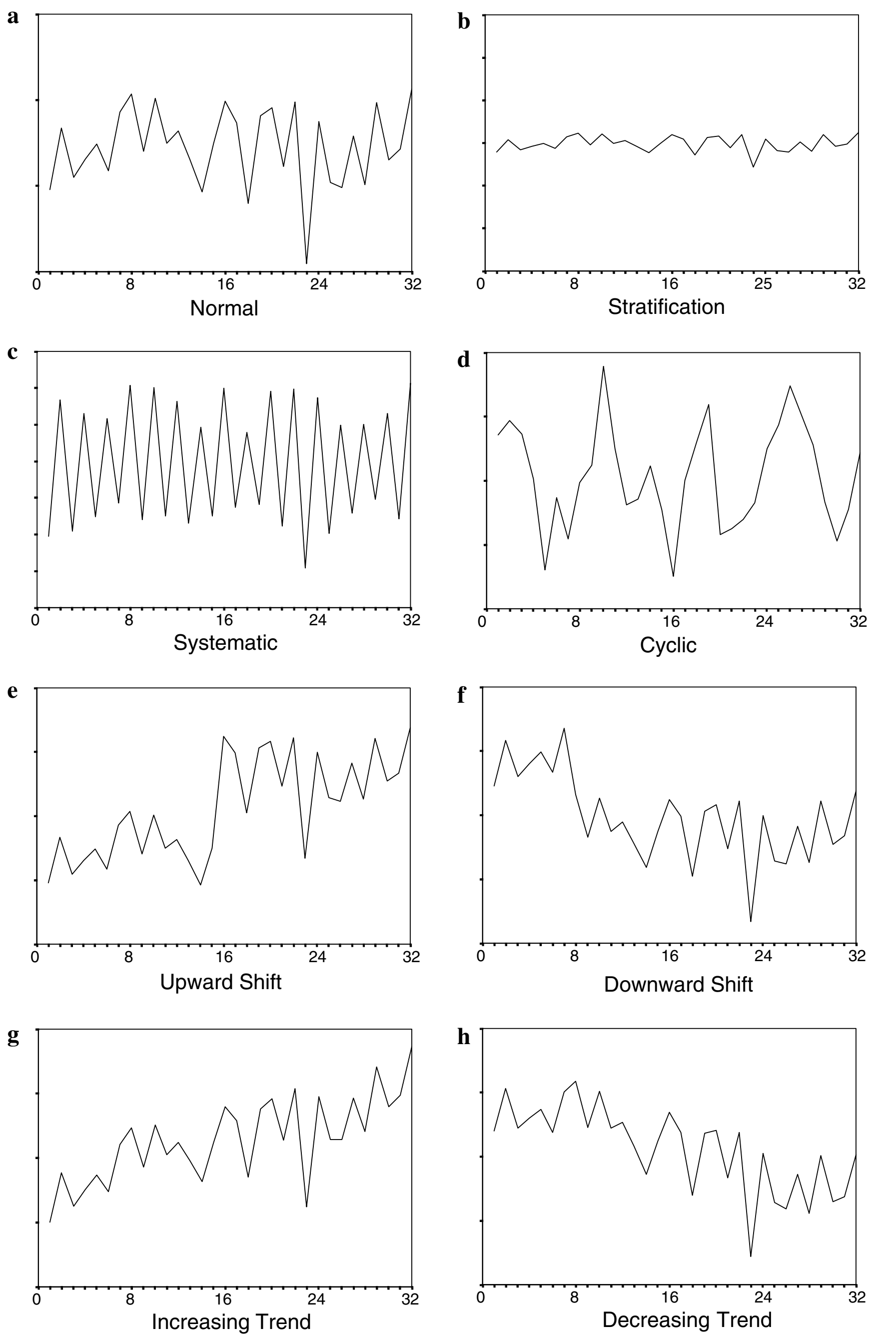


Fig. 1. Various control chart patterns.

control. An unnatural pattern identifies a process when it is out of control. The unnatural patterns are associated with impending problems having assignable causes requiring pre-emptive actions (Western Electric, 1958). Identification of various types of unnatural patterns can greatly narrow down the set of possible causes that must be investigated and thus the diagnostic search process can be effectively reduced in length.

However, recognition of unnatural patterns is a critical task in statistical process control (SPC). Over the years, numerous supplementary rules known as zone tests or run tests (Grant & Leavenworth, 1996; Nelson, 1984; Nelson, 1985; Western Electric, 1958) have been suggested to the practitioners to detect unnatural patterns. Application of all these rules can result in excessive number of false alarms. Nevertheless, identification and analysis of the unnatural patterns require considerable experience and skill from the part of the practitioners. Since actions can be taken promptly if the operators can themselves detect the out-of-control situations, it is desirable that the shop floor people implement the control charts. However, usually they are lacking the skill and experience that are needed for the interpretation of control chart patterns.

An efficient automated control chart pattern (CCP) recognition system can compensate this gap and ensure consistent and unbiased interpretation of CCPs leading to lesser number of false alarms and easier implementation of control charts. A number of works aiming this have been reported in the literatures. Most of the researchers (Al-Assaf, 2004; Assaleh & Al-Assaf, 2005; Guh & Tannock, 1999; Hassan, Nabi Baksh, Shaharoun, & Jamaluddin, 2003; Pham & Wani, 1997; Pham & Sagiroglu, 2001; Yang & Yang, 2002) have been concerned with the recognition of normal, cyclic, upward shift, downward shift, increasing trend and decreasing trend patterns. Other researchers (Cheng, 1997; Guh, Zorriassatine, Tannock, & O'Brien, 1999a, Guh, Tannock, & O'Brien, 1999b; Guh & Shiue, 2005; Perry, Spoerre, & Velasco, 2001; Wang, Kochhar, & Hannam, 1998; Yang & Yang, 2005) have taken into consideration some other unnatural patterns like systematic and/or mixture also. However, many real life situations demand for a CCP recognition system that can be capable of detecting stratification pattern too. A stratification pattern differs from a normal pattern due to the location of the observations in relation to the centerline and the control limits. In other words, a stratification pattern may be viewed as a normal pattern with unexpectedly lower variability. Majority of the reported research works have been carried out on scaled pattern data. The distinction between normal and stratification patterns is lost when the observations are scaled to $[0,1]$ or $[-1,1]$ interval. This may be the possible reason why most of the research works do not take into account the stratification pattern. Only Hwarng and Hubele (1993) have attempted to recognize stratification pattern and their work is based on standardized pattern data.

The approaches adopted by the researchers for developing a CCP recognition system have ranged from the application of expert systems (Evans & Lindsay, 1988; Pham & Oztemel, 1992a; Swift & Mize, 1995) to neural networks (Cheng, 1997; Guh & Tannock, 1999; Guh et al., 1999a, 1999b; Guh & Shiue, 2005; Hwarng & Hubele, 1993; Pham & Oztemel, 1992b; Pham & Sagiroglu, 2001; Pham & Oztemel, 1994; Perry et al., 2001; Velasco & Rowe, 1993; Yang & Yang, 2002) to feature-based heuristics/neural networks (Hassan et al., 2003; Pham & Wani, 1997). The advantage of an expert system or rule-based system is that it contains the information explicitly. If required, the rules can be modified and updated easily. However, the use of rules based on statistical properties has the difficulty that similar statistical properties may be derived for some patterns of different classes, which may create problems of incorrect recognition. The advantage with neural network is that it does not require the provision of explicit rules or templates. Rather, it learns to recognize patterns directly through typical example patterns during a training phase. A neural network also has the ability to identify an arbitrary pattern not previously encountered. However, there is no guarantee that it will identify such patterns correctly. One disadvantage with neural network is that the information it contains, is implicit and virtually inaccessible to the user. This creates difficulties in understanding how a particular classification decision has been reached and also in determining the details of how a given pattern resembles with a particular class. In addition, there is no systematic way to select the topology and structure of a neural network. In general, this has to be found empirically, which can be time consuming. The feature-based approach has two main steps, i.e., (a) extraction of features and (b) recognition of CCPs using those extracted features. The CCPs can be recognized directly from the extracted features based on properly defined heuristics or a neural network can be trained for CCP recognition using those extracted features as the input vector. The feature-based neural network approach reduces network size and learning time. Since extracted shape features represent the main characteristics of the original data in a condensed form, both the feature-based heuristic and feature-based neural network approaches can facilitate efficient pattern recognition. However, the feature-based heuristic approach has the distinct advantage that the practitioners can clearly understand how a particular pattern is identified by the use of relevant shape features, which is very important in gaining their confidence in CCP recognition. This, in turn, increases chances of successful implementation of control charts.

Pham and Wani (1997) and Hassan et al. (2003) have demonstrated the usefulness of feature-based neural network approach in discriminating six types of patterns, e.g., normal, cyclic, upward shift, downward shift, increasing trend and decreasing trend. Hassan et al. (2003) have utilized six statistical features, whereas Pham and Wani (1997) have used nine shape features for the pattern recognition task. One limitation in the extraction of statistical features is that it requires considerably large number of observations. Moreover, the statistical features lose information on the order of the data. On the other hand, each type of pattern has its own geometric shape and various related features can represent this shape. The advantage of shape features is that it can be extracted from lesser number of observations without losing order of the data. Pham and Wani (1997) have demonstrated that properly developed heuristics based on the extracted shape features can also efficiently differentiate various CCPs. However, their CCP recognition system is not truly automated, since extraction of some of their proposed features requires user's inputs, necessitating the utilization of skill and experience of the end-users. Moreover, the magnitudes of some of their proposed features (if extracted from the actual data) are dependent on the inherent process variation and also on the available number of observations, and thus generalized interpretation of those features becomes difficult. The aim of this work is to explore the possibilities of developing a truly automated feature-based CCP recognition system that can differentiate all the eight commonly observed CCPs, i.e., normal, stratification, systematic, cyclic, upward shift, downward shift, increasing trend and decreasing trend patterns. Two types of CCP recognizers using feature-based heuristic and feature-based neural network approaches are developed and their relative performances are extensively studied.

This paper is organized in different sections. A brief review of the shape features proposed by Pham and Wani (1997) is provided in Section 2. The sample patterns and their simulation methodologies are discussed in Section 3. Eight new shape features, extraction of which do not require inputs from the practitioners in any form, are proposed and presented in Section 4. Section 5 discusses design of the pattern recognizers followed by the experimental procedures in Section 6. Section 7 provides the results and discussions on the comparison between two types of recognizers. Section 8 presents some conclusions.

## 2. Review of shape features proposed by Pham and Wani

Pham and Wani (1997) have considered nine features for discriminating six control chart patterns. The features are as follows:

(a) Slope ($s$) of the least square line representing the pattern,
(b) number of mean crossings (nc1),
(c) number of least square line crossings (nc2),
(d) area between the pattern and the mean line ($a_{\mathrm{pm}}$),
(e) area between the pattern and its least square line ($a_{\mathrm{ps}}$),
(f) cyclic membership (cmember),
(g) average slope of the line segments (as),
(h) slope difference (sd), and
(i) area between the least square line and the line segments ($a_{\mathrm{ss}}$).

The computations of the first five features are straightforward. The values of as, sd and $a_{\mathrm{ss}}$ are computed by dividing the observation window into two segments and fitting two least square lines to the observations in these two segments. The discriminatory power of these features depends on how good is the segmentation. The authors left the task of segmentation to the users and allowed that a window size could be as low as five points. The need for user's inputs calls for utilizing his experience and skill, which is undesirable and against the very purpose of the CCP recognizer. To avoid choosing incorrect limit points for segmentation, the authors have proposed imposition of a constraint that the line segments obtained would not cross the mean line. However segmentation satisfying the above constraint may not be feasible under some sort of random fluctuations (see Fig. 2) and thus the three features (as, sd and $a_{\mathrm{ss}}$) that are supposed to discriminate trend from shift patterns, become inestimable. It may be worth to mention here that an attempt of segmentation using the constraint of Pham and Wani on 1200 simulated patterns (see Section 3) reveals that the segmentation is infeasible for about 12% of the patterns.

On the other hand, the feature cmember is considered based on the fact that for a cyclic pattern, the half period (interval between two consecutive mean crossings) is nearly constant and the area between the pattern and the mean line is similar for each half period. The cyclic pattern is expected to be segregated from the

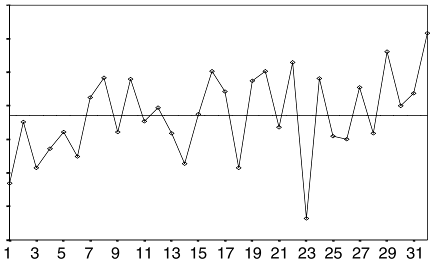


Fig. 2. A pattern, where segmentation satisfying the constraint of Pham and Wani, is infeasible.

normal pattern by the value of cmember. For the calculation of cmember value, the average area ($\bar{a}$) and average distance ($\bar{d}$) of those half cycles in a pattern which have a period of greater than or equal to the minimum possible period of the cyclic pattern are estimated. Then starting from the first mean line crossing, area $a'$ and distance $d'$ of each half-cycle are obtained and cmember value is calculated by comparing them with the ranges $\bar{a} \pm \Delta a$ and $\bar{d} \pm \Delta d$, respectively. It may be noted that $\Delta a$ and $\Delta d$ values are assigned by the user and thus the procedure for estimation of cyclic membership value again requires utilization of the experience and knowledge of the user, which is contradictory to the very purpose of the CCP recognizer. Furthermore, when the period and amplitude of cyclic pattern are less, resemblance of normal pattern with cyclic pattern is not uncommon (see Fig. 3) and consequently cyclic pattern becomes undistinguishable from normal pattern based on cyclic membership value. Comparison of cmember values for simulated normal and cyclic patterns with periodicity $(T) = 8$ and amplitude $(a) = 1.5\sigma$, $2.0\sigma$ and $2.5\sigma$ (all generated from a set of 32 time series of standard normal data with 32 values in time series; $\sigma$ is the process standard deviation) reveals that cyclic patterns cannot be differentiated well from normal patterns when the amplitude of cyclic pattern is less than $1.5\sigma$ (see Fig. 4). Therefore cmember cannot be a good feature for consideration.

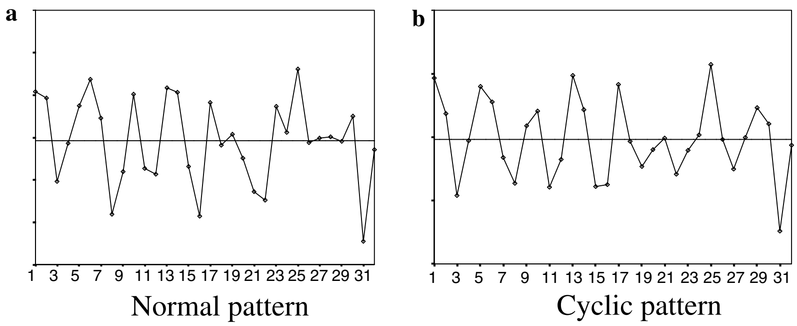


Fig. 3. A normal pattern and a cyclic pattern that is undistinguishable.

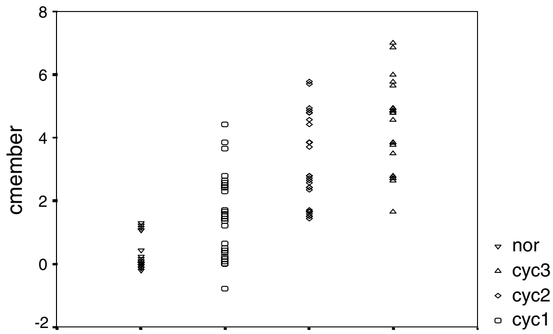


Fig. 4. cmember values for normal and cyclic patterns with different amplitudes [cyc1($a = 1.5\sigma$), cyc2($a = 2.0\sigma$) and cyc3($a = 2.5\sigma$)].

Another important issue is the difficulty in generalized interpretation about the values of some of the features. For example, the values of nc1 and nc2 for normal patterns with 32 and 64 observations will be different. Better definitions of these features will be the percentage of mean line crossover and percentage of least square line crossover over the total number of observations, respectively, so that the threshold values of these features for different patterns become insensitive to the number of available observations. On the other hand, the values of $a_{pm}$,$a_{ps}$ and $a_{ss}$ are highly dependent on process variation (if extracted from actual data). A slight change in the process variation will lead to changes in the threshold values of these features for different patterns resulting in higher misclassification error.

## 3. Sample patterns

Sample patterns are required for assessing the usefulness of various possible shape features in discriminating different patterns as well as developing/validating a CCP recognizer. Ideally, sample patterns should be collected from a real manufacturing process. Since, a large number of patterns are required for developing and validating a CCP recognizer and as those are not economically available, simulated data are often used. This is a common approach adopted by other researchers as mentioned in the earlier section.

Pham and Wani (1997) have considered an observation window with 64 data points. A large window size can lower the recognition efficiency by increasing the time required to detect the patterns. In this study, an observation window with 32 data points is considered implying that each sample pattern consists of 32 time-sequenced data. Most of the researchers (Cheng, 1997; Guh & Tannock, 1999; Guh et al., 1999a, 1999b; Guh & Shiue, 2005; Hwarng & Hubele, 1993; Hassan et al., 2003; Yang & Yang, 2005) have assumed that only one fundamental period exists for cyclic patterns, and/or a sudden shift appears only in the middle of an observation window. These are very restrictive assumptions. In this study, existence of multiple periods for cyclic patterns and appearance of sudden shift randomly at any one of the three time points between one-fourth and three-fourth observations are considered. The parameters for simulating the eight main CCPs are given in Table 1. The values of different parameters for unnatural patterns are varied randomly in a uniform manner between the limits shown. The minimum parameter values are chosen such that the patterns can be sufficiently differentiable even after being contaminated by random variations. The equations used for simulating the eight CCPs are given in Appendix A. A total of 1200 ($150 \times 8$) and 1000 ($125 \times 8$) sample patterns are used in the training and validation/verification phases, respectively.

## 4. Proposed pattern features

A set of eight new shape features is considered which can segregate well all the eight commonly observed CCPs. As mentioned in Section 1, a stratification pattern differs from a normal pattern due to the location of the observations in relation to the centerline and control limits. In other words, a stratification pattern may be viewed as a normal pattern with unexpectedly lower variability. The distinction between normal and

Table 1
Parameters for simulating control chart patterns

| Pattern type | Parameters | Values |
|---|---|---|
| Normal (NOR) | Mean ($\mu$) | 80 |
| | Standard deviation ($\sigma$) | 5 |
| Stratification (STA) | Random noise ($\sigma'$) | $0.2\sigma$ to $0.4\sigma$ |
| Systematic (SYS) | Systematic departure ($d$) | $1\sigma$ to $3\sigma$ |
| Cyclic (CYC) | Amplitude ($a$) | $1.5\sigma$ to $2.5\sigma$ |
| | Period ($T$) | 8 and 16 |
| Increasing trend (UT) | Gradient ($g$) | $0.05\sigma$ to $0.1\sigma$ |
| Decreasing trend (DT) | Gradient ($g$) | $-0.1\sigma$ to $-0.05\sigma$ |
| Upward shift (US) | Shift magnitude ($s$) | $1.5\sigma$ to $2.5\sigma$ |
| | Point of shift ($P$) | 9th, 17th, 25th observations |
| Downward shift (DS) | Shift magnitude ($s$) | $-2.5\sigma$ to $-1.5\sigma$ |
| | Point of shift ($P$) | 9th, 17th, 25th observations |

stratification patterns is lost when the observations are scaled to $[0,1]$ or $[-1,1]$ interval. Extraction of various shape features from the data plot of actual observations can alleviate this problem. In this paper, therefore, all the features are extracted from the control chart plot of actual observations assuming that each interval in the control chart is represented by 1 U in linear measurement scale. Extraction of features from the data plot of actual observations has the advantage that interpretation for the values of various features in relation to the control chart plot becomes simpler and thus their acceptability to the practitioners will be higher psychologically. It may be noted that some of the proposed features described below are expressed in terms of the standard deviation. It is so done in order to make those features least sensitive to changes in the process variation. On the other hand, all the proposed features are defined in such a way that their magnitudes are independent of mean of the underlying process.

Among the eight features, extraction of four features does not require segmentation of the data plot, but the other four require segmentation. Two segmentation criteria are proposed. One is the predetermined segmentation into four windows of equal size and the other is segmentation into two windows such that the pooled mean sum of squares of error (PMSE) of the two least square (LS) lines fitted to these windows will be the minimum.

### 4.1. Features extracted without segmentation of observations

(a) Absolute slope of the LS line representing the pattern (AB):

$$\mathrm{AB} = \left| \frac{\sum_{i=1}^{N} y_i (t_i - \bar{t})}{\sum_{i=1}^{N} (t_i - \bar{t})^2} \right| \tag{1}$$

where $t_i(=i)$ $(i = 1, 2, \ldots, N)$ is the distance of $i$th time point of observation from the origin, $y_i$ is the observed value of a quality characteristic at $i$th time point, $N$ is the total number of observations and $\bar{t} = \frac{\sum_{i=1}^{N} t_i}{N}$.

The magnitude of AB for trend and shift patterns is greater than zero. For all other patterns, i.e., normal, stratification, systematic and cyclic, the slope is around zero. Therefore, the magnitude of the absolute slope of the LS line can differentiate the four patterns that hang around the centerline, i.e., normal, stratification, systematic and cyclic patterns from trend and shift patterns.

(b) Sign of the slope of the LS line representing the pattern (SB): the SB can be represented as a categorical variable, which is '0' if the sign is negative and '1' otherwise. It can make discrimination between decreasing and increasing trend patterns, and downward and upward shift patterns.

It may be noted that the slope of LS line ($s$) alone as a feature can give equivalent results that may be derived from the two features AB and SB. But considering slope of LS line ($s$) as a single feature may lead to unreasonable split in '$s$' (say, $s < -0.005$ and $s \geqslant -0.005$) to differentiate decreasing trend or downward shift patterns from increasing trend or upward shift patterns when a set of learning samples is subjected to analysis in a classification tree program (discussed in Section 5.1). Therefore, the two features AB and SB are considered instead of '$s$'.

(c) Average distance (RDIST) between the consecutive points expressed in terms of SD:

$$\mathrm{RDIST} = \left( \frac{1}{N-1} \left[ \sum_{i=1}^{N-1} \sqrt{(t_{i+1} - t_i)^2 + (y_{i+1} - y_i)^2} \right] \right) \Big/ \mathrm{SD}; \quad \mathrm{SD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{2}$$

Analysis of the synthetic pattern data reveals that the magnitude of RDIST is higher for stratification and systematic patterns, intermediate for normal pattern and on the lower side for all other patterns (see Fig. 5). Thus the magnitude of RDIST can differentiate stratification and systematic patterns from normal and cyclic patterns and also shift and trend patterns from normal pattern.

(d) Area between the pattern and LS line per LS line crossover expressed in terms of SD (ALSLSC):

$$\mathrm{ALSLSC} = (\mathrm{ALS}/\mathrm{LSC})/\mathrm{SD} \tag{3}$$

where, ALS is the area between the pattern and the LS line fitted to the observations and LSC is the number of LS line crossover (see Pham & Wani, 1997; for graphical interpretations of ALS and LSC). The value of ALS can be computed easily by summing the areas of the triangles and trapeziums that are formed by the LS line and overall pattern.

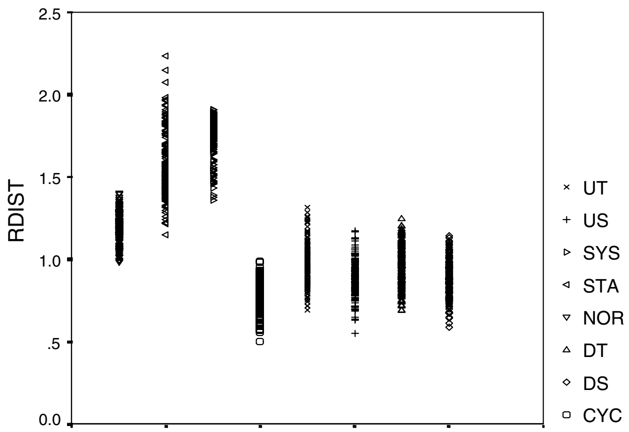


Fig. 5. Values of RDIST for different patterns.

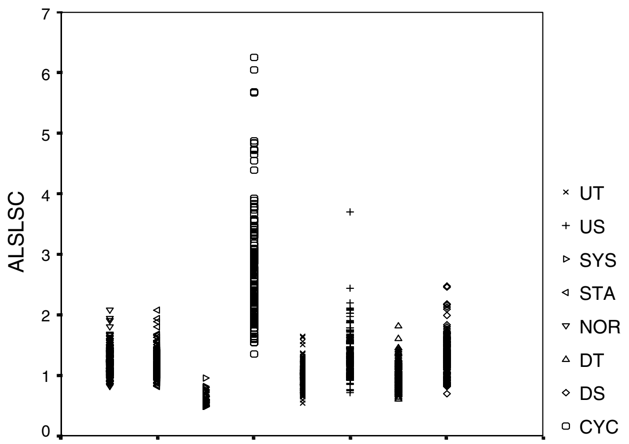


Fig. 6. Values of ALSLSC for different patterns.

Among the four patterns that are built surrounding the mean line, ALSLSC is the maximum for cyclic pattern, intermediate for normal and stratification patterns and least for systematic pattern (see Fig. 6). Therefore the magnitude of ALSLSC can differentiate cyclic and systematic patterns from normal and stratification patterns. It can also differentiate shift patterns from trend patterns.

### 4.2. Features extracted after segmentation of observations

One of the most difficult tasks in control chart pattern recognition is to differentiate shift patterns from trend patterns. It is not unlikely that a shift occurred at a random point under some random fluctuation of observations resembles a trend pattern occurred under a different random fluctuation of observations. Therefore analyses of control chart patterns are required from multiple angles for differentiating shift patterns from trend patterns. Instead of judgment-based segmentation by the user (as proposed by Pham & Wani, 1997), two different segmentation criteria are considered, i.e., (a) pre-determined segmentation, where the window sizes are fixed and (b) criterion-based segmentation, where the window sizes may vary in order to satisfy the desired criterion.

#### 4.2.1. Predefined segmentation of observations

In this approach, the observation window is divided into four segments having equal size, as shown in Fig. 7. Then the features are extracted based on fitting of LS lines to different subsets of $N/2$ observations.

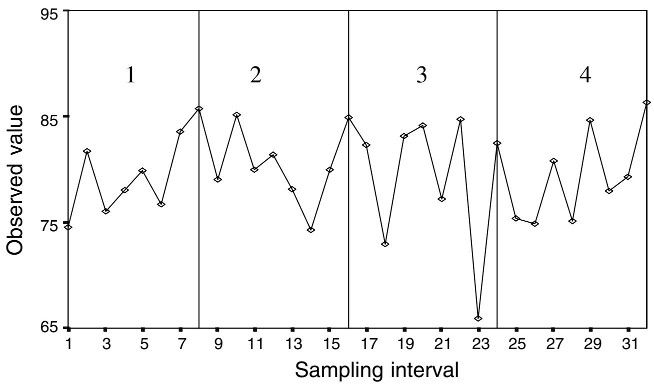


Fig. 7. Four segments of equal size in a pattern.

A subset of $N/2$ observations can be obtained by combining any two segments in $C_2^4 = 6$ ways. A shift can occur in any of the segments 2, 3 and 4. In case of a shift pattern, the slopes of the least square lines fitted to the subsets of $N/2$ observations will be widely varying. For some subsets of observations, the slopes will be almost zero and for other subsets, the slopes will be very steep (see Fig. 8). The same will happen in case of a cyclic pattern unless each segment contains a complete cycle. Whereas in case of a trend pattern, the slopes of least square lines fitted to all the subsets of $N/2$ observations will be similar but different from zero. For the remaining patterns, slopes of the LS lines fitted to all the subsets of observations are expected to be about zero. Taking this fact into consideration, the following feature is defined.

Range of slopes of LS lines fitted to six subsets of $N/2$ observations (BRANGE):

$$\mathrm{BRANGE} = \mathrm{Maximum}(b_{jk}) - \mathrm{Minimum}(b_{jk}); \quad (j = 1, 2, 3;\ k = 2, 3, 4;\ j < k) \tag{4}$$

where, $b_{jk}$ is the slope of the LS line fitted to the observations in $j$th and $k$th segments.

The magnitude of BRANGE will be higher for shift patterns than trend patterns. Unless each segment contains a complete cycle, the magnitude of BRANGE will also be higher for cyclic pattern than normal, stratification and systematic patterns. Therefore, it can differentiate shift patterns from trend patterns, and cyclic pattern from normal, stratification and systematic patterns.

### 4.2.2. Criterion-based segmentation

It is well-known that the straight line that can be fitted best to a given set of observations is the least square (LS) line. In case of a shift pattern, the total number of observations can be divided into two segments (before and after the occurrence of a shift) and two LS lines (each approximately horizontal to the $X$-axis) can be fitted well in these segments. However, the time point of occurrence of the shift cannot be known exactly. Therefore, a criterion-based segmentation into two segments, where the segment sizes may vary in order to satisfy the desired criterion, is proposed. The two LS lines that lead to the minimum PMSE may be considered as the two best-fitted LS lines within the overall pattern. The PMSE value of the two LS lines is given by the following expression:

$$\mathrm{PMSE} = \frac{\left[\sum_{j=1}^{2} \mathrm{df}_j \times \mathrm{MSE}_j\right]}{\sum_{j=1}^{2} \mathrm{df}_j}; \quad (j = 1, 2) \tag{5}$$

where, $\mathrm{MSE}_j$ and $\mathrm{df}_j$ are the mean sum of squares of errors for the fitted LS line and its degrees of freedom in $j$th window, respectively. Let the number of observations in first and second windows are $N_1$ and $N_2$, respectively. Assuming that at least eight observations are needed for fitting an LS line ($8 \leqslant N_1 \leqslant 24$, $N_1 + N_2 = N$), LS lines are fitted to all possible two windows and PMSE values are computed. Then the value of $N_1$ that leads to the minimum PMSE is considered as the limit point of the first window and $(N_1 + 1)$ becomes the starting point of the second window. The following three features are extracted after segmentation that will minimize the PMSE of the LS lines fitted to two windows.

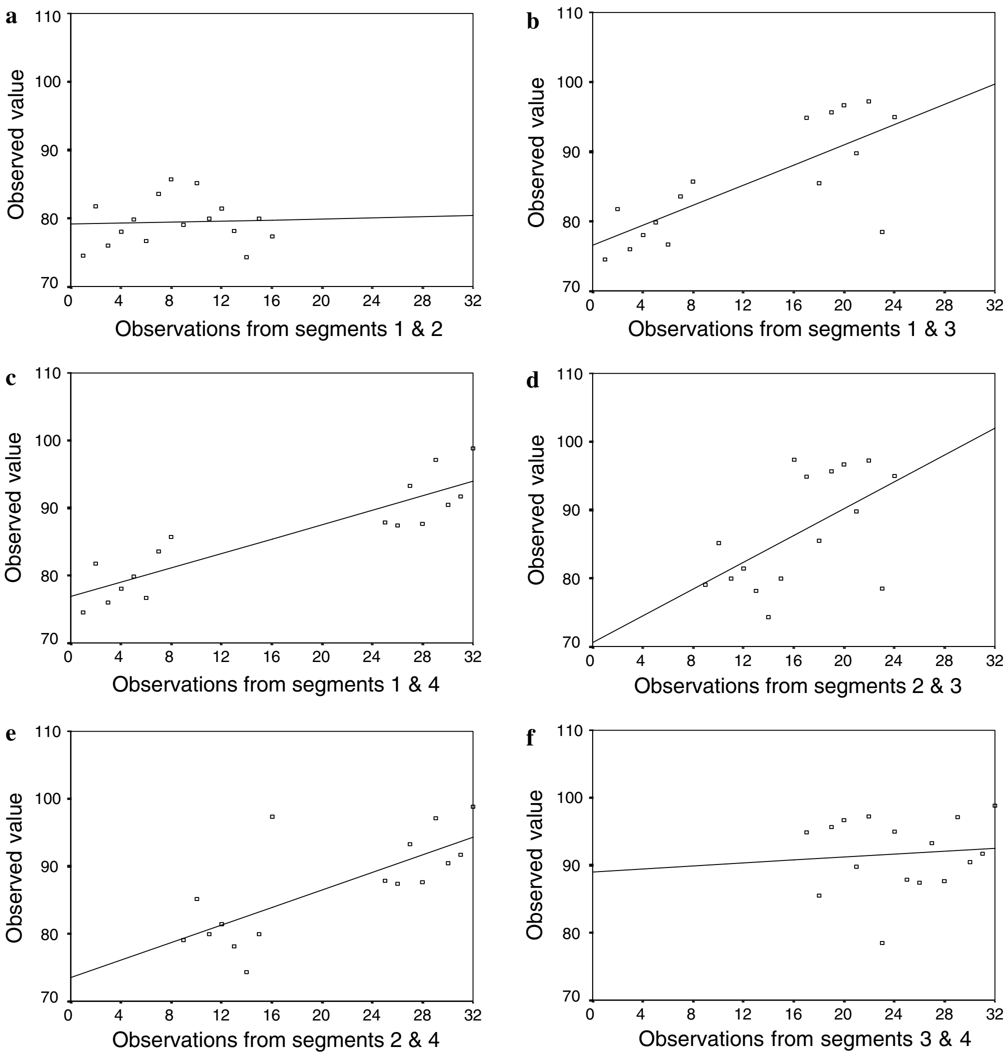


Fig. 8. Slopes of LS lines fitted to six subsets of $N/2$ observations in a shift pattern.

(a) Absolute average slope of the two LS lines (AABPE):

$$\mathrm{AABPE} = \left| \frac{1}{2} \sum_{j=1}^{2} B_j \right|; \quad (j = 1, 2) \tag{6}$$

where, $B_j$ is the slope of the LS line fitted to $j$th window when the segmentation minimizes PMSE of the LS lines fitted to two windows.

The value of AABPE for trend patterns will be higher than shift patterns. On the other hand, among the four patterns that are built surrounding the center line, AABPE is the highest for cyclic pattern. Therefore, the magnitude of AABPE can differentiate trend patterns from shift patterns, and cyclic pattern from normal, systematic and stratification patterns.

(b) Absolute slope difference between the LS line representing the overall pattern and the average slope of LS lines representing the patterns within the two segments (ABDIFPE):

$$\mathrm{ABDIFPE} = \left| B - \left( \frac{1}{2} \sum_{j=1}^{2} B_j \right) \right|; \quad (j = 1, 2) \tag{7}$$

where, $B$ is the slope of the LS line representing the overall pattern.

The magnitude of ABDIFPE will be higher for shift patterns than trend patterns. On the other hand, among the four patterns that are hang around the mean line, the magnitude of ABDIFPE is the highest for cyclic pattern followed by systematic pattern, intermediate for normal pattern and least for stratification pattern. This feature, therefore, can differentiate shift patterns from trend patterns, and normal pattern from cyclic, systematic and stratification patterns.

(c) Ratio of MSE of the LS line representing the overall pattern and PMSE of the LS lines fitted to two windows (REPEPE):

$$\mathrm{REPEPE} = \mathrm{MSE}/\mathrm{PMSE} \tag{8}$$

The magnitude of REPEPE will be higher for shift and cyclic patterns and lesser for all other patterns. Therefore, it can discriminate shift patterns from trend patterns, and cyclic pattern from normal, stratification and systematic patterns.

It may be noted that the usefulness of these three features AABPE, ABDIFPE and REPEPE are similar to the features $a_{\mathrm{pm}}$, $a_{\mathrm{ps}}$ and $a_{\mathrm{ss}}$ (considered by Pham & Wani, 1997), respectively. However, the above three features are always estimable, whereas $a_{\mathrm{pm}}$, $a_{\mathrm{ps}}$ and $a_{\mathrm{ss}}$ are not so due to the reasons illustrated in Section 2.

## 5. Pattern recognizer design

It is evident from the discussions in Section 1 that the feature-based approach for CCP recognition can be of two types, i.e., (a) feature-based heuristics and (b) feature-based neural network. Both these approaches have their own merits and demerits. On the other hand, an artificial neural network (ANN) recognizer can also be developed that will use the raw data as the input vector. According to the reported results (Hassan et al., 2003; Pham & Wani, 1997), an ANN recognizer with raw data as the input vector gives poorer recognition performance and so this type of ANN recognizer is not considered here.

### 5.1. Feature-based heuristics

This technique uses simple IF ...(condition)....THEN ....(action)..... heuristic rules. The conditions for the rules set the threshold values of the features and the actions are the classification decisions. The set of heuristics, arranged as a decision tree, provides easily understood and interpreted information regarding the predictive structure of the data for various features. However, determination of the optimal set of heuristic rules is a critical issue. Pham and Wani (1997) have derived the classification tree by manually inspecting the feature values in a set of learning sample. However manual process of obtaining a good set of heuristics is extremely laborious and thus it is impracticable. An alternative and preferred approach is to make usage of the tree-structured classification algorithms so that the burden of manual process is eliminated, time is saved and hence, accuracy is improved.

A variety of classification tree programs have been developed to predict the membership of cases or objects in the class of a categorical dependent variable (pattern class) from their measurements on one or more predictor variables (features). Some classification tree programs, such as FACT (Loh & Vanichestakul, 1988) and THAID (Morgan & Messenger, 1973) perform multilevel splits while computing the classification trees. Some other classification tree programs, such as CART (Classification and Regression Trees) (Breiman, Friedman, Olshen, & Stone, 1984) and QUEST (Quick, Unbiased, Efficient Statistical Trees) (Loh & Shih, 1997) perform binary splits when developing the classification trees. A multilevel split divides a parent node into more than two child nodes, but a binary split always produces just two nodes (regardless of the number of levels of the splitting variable or the number of classes on the dependent variable). It may be noted that there is no inherent advantage of multilevel splits, because any multilevel split can be represented as a series of binary splits. In some programs that perform multilevel splits, the predictor variables can be used for splitting only once, so the resulting classification trees may be unrealistically short and the most serious problem is the bias in variable selection for splits.

Since multilevel split does not provide any inherent advantage and interpretation is relatively easier for binary classification tree, use of CART or QUEST algorithm is preferable for developing the tree. From a preliminary study, it is found that the classification tree resulting from QUEST algorithm, in general, leads to higher misclassification error. It is planned, therefore, to adopt CART algorithm for developing feature-based heuristics.

The CART algorithm uses an exhaustive grid search of all possible univariate splits of the predictor variables (features) to find out the optimal split for a classification tree. As differential base rates are reflected in the class sizes, prior probabilities are estimated by the class proportions in the sample. The CART program allows automatic selection of the 'right-sized' tree that has the optimal predictive accuracy. The procedures for the 'right-sized' tree selection are not foolproof, but at least, they take the subjective judgment out of the process of selecting the 'right-sized' tree and thus avoid 'over fitting' and 'under fitting' of the data. The split conditions and size of the classification tree are determined based on the specifications of the following parameters:

Misclassification cost of a pattern: misclassification cost for different patterns may be different or equal.
Measure of goodness of fit for a split: at each node, the split that generates the greatest improvement in the predictive accuracy is selected. This is measured with some type of node impurity measure like Gini measure, Chi-square measure, G-square measure.
The value of '*n*' for 'Minimum *n*' rule: the 'Minimum *n*' rule implies that splitting is allowed to continue until all terminal nodes are pure or contain no more than '*n*' cases.
Pruning (recombining): pruning is performed to produce a sequence of optimally pruned and nested trees, which facilitate selection of the 'right sized' tree. There are two types of pruning, i.e., (a) minimal cost-complexity cross-validation (CV) pruning (Breiman et al., 1984) and (b) minimal deviance-complexity cross-validation pruning (Ripley, 1996).
The value of $\delta$ for '$\delta$ Standard Error' rule: the '$\delta$ Standard Error' rule implies choosing the smallest-sized tree from all trees whose CV cost does not exceed the minimum CV cost plus $\delta$ times the standard error of the minimum CV cost as the right sized tree.

### 5.2. Feature-based neural network

A multiplayer perceptrons (MLP) is used as pattern recognizer. The MLP architecture has been successfully applied to solve some difficult and diverse problems in modeling, prediction and pattern classification (Haykin, 1999). This type of neural network is simple and ideally suited for pattern recognition tasks. Its basic structure comprises an input layer, one or more hidden layer(s) and an output layer. The number of input nodes in the input layer is set according to the actual number of features used, i.e., eight. The number of output nodes in this study is set corresponding to the number of pattern classes, i.e., eight. The net has used one hidden layer and number of nodes in this hidden layer is set to 12 (chosen empirically). Thus the ANN structure is $8 \times 12 \times 8$.

Since it is planned to use the supervised training approach, each pattern presentation is tagged with its respective label. The labels, shown in Table 2, are the target values for the recognizers' output nodes. The maximum value in each row (0.9) identifies the corresponding node expected to secure the highest output for a pattern to be considered correctly classified.

The back propagation training algorithm, gradient descent with momentum and adaptive learning rate (*traingdx*) is adopted here for training the network, since it provides reasonably good performance and more

Table 2
Target recognizer outputs

| Pattern class | Description | Target recognizer outputs Node | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | NOR | 0.9 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| 2 | STA | 0.1 | 0.9 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| 3 | SYS | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| 4 | CYC | 0.1 | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 | 0.1 | 0.1 |
| 5 | UT | 0.1 | 0.1 | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 | 0.1 |
| 6 | US | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.9 | 0.1 | 0.1 |
| 7 | DT | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.9 | 0.1 |
| 8 | DS | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.9 |

consistent results (Hassan et al., 2003). It is also more memory-efficient. The activation functions used are hyperbolic tangent (*tansig*) for the hidden layer and sigmoid (*logsig*) for the output layer. The hyperbolic tangent function transforms the layer inputs to output range from −1 to +1 and the sigmoid function transforms the layer inputs to output range from 0 to 1 (Smith, 1993).

## 6. Experimental procedure

Six sets of training samples of size 1200 each are generated. Each set of the training sample is then subjected to CART analysis using the following specifications:

Misclassification cost of a pattern – equal for all the patterns;
Measure of goodness of fit for a split – Gini measure;
Value of '$n$' for 'Minimum $n$' rule = 1;
Pruning (recombining) – minimal cost-complexity cross-validation pruning;
Value of '$\delta$' for '$\delta$ standard error' rule = 0.3.

This results six different classification trees, i.e., six heuristic CCP recognizers. These recognizers are labeled 1.1–1.6 in Table 3.

On the other hand, the neural network is trained six times by exposing it separately to the six sets of training samples with the following training parameters:

Maximum number of epochs = 1500;
Error goal = 0.01;
Learning rate = 0.5;
Momentum constant = 0.5;
Ratio to increase learning rate = 1.05;
Ratio to decrease learning rate = 0.7.

The training is stopped whenever either the error goal has been achieved or the maximum allowable number of training epochs has been met. In this process, six different feature-based ANN recognizers are developed. All these six ANN recognizers have the same architecture and differ only in the training data sets used. These recognizers are labeled 2.1–2.6 in Table 4.

The recognition performances of all the six feature-based heuristic recognizers as well as feature-based ANN recognizers are tested (verified) using six different data sets of size 1000 each. Results of this verification are presented and discussed below.

## 7. Results and discussion

This section presents the results and comparisons of the performance between the feature-based heuristic recognizers and ANN recognizers. Tables 3 and 4 show the training and verification performance of the six

Table 3
Training and verification performance for feature-based heuristic recognizers

| Recognizer number | Training phase | | Verification phase | | | |
|---|---|---|---|---|---|---|
| | Number of splits in the tree | Percentage correct classification | Percentage correct classification | | | |
| | | | Mean | Max. | Min. | Range |
| 1.1 | 18 | 96.1 | 95.25 | 98.1 | 94.2 | 3.9 |
| 1.2 | 22 | 96.8 | 94.94 | 97.1 | 94.2 | 2.9 |
| 1.3 | 24 | 97.2 | 95.22 | 96.4 | 94.0 | 2.4 |
| 1.4 | 17 | 95.9 | 94.68 | 96.8 | 93.4 | 3.4 |
| 1.5 | 22 | 95.2 | 94.74 | 96.2 | 93.0 | 3.2 |
| 1.6 | 21 | 96.4 | 94.84 | 96.6 | 93.4 | 3.2 |
| Overall mean | 20.6 | 96.26 | 94.95 | Overall range | | 5.1 |

Table 4
Training and verification performance for feature-based ANN recognizers

| Recognizer number | Training phase | | Verification phase | | | |
|---|---|---|---|---|---|---|
| | Number of epochs | Percentage correct classification | Percentage correct classification | | | |
| | | | Mean | Max. | Min. | Range |
| 2.1 | 1001 | 96.2 | 96.37 | 96.9 | 95.8 | 1.1 |
| 2.2 | 518 | 96.9 | 96.18 | 96.9 | 95.2 | 1.7 |
| 2.3 | 589 | 95.8 | 95.77 | 96.6 | 94.8 | 1.8 |
| 2.4 | 613 | 96.0 | 96.15 | 96.8 | 95.1 | 1.7 |
| 2.5 | 1113 | 95.9 | 95.88 | 97.0 | 94.7 | 2.3 |
| 2.6 | 498 | 96.3 | 96.39 | 97.2 | 95.9 | 1.3 |
| Overall mean | 722 | 96.19 | 96.13 | Overall range | | 2.5 |

feature-based heuristic recognizers and six feature-based ANN recognizers, respectively. The overall mean percentages of correct recognition for feature-based heuristic recognizers and ANN recognizers at training phase are quite similar, 96.26% and 96.19%, respectively. On the other hand, the overall mean percentages of correct recognition for feature-based heuristic recognizers and ANN recognizers at verification phase are 94.95% and 96.13%, respectively. These percentage values range from 94.68% to 95.25% and 95.77% to 96.39% for the heuristic and ANN recognizers, respectively. This indicates that the feature-based ANN recognizers have better recognition performance than feature-based heuristic recognizers. On the other hand, the overall range of percentage values of correct recognition at the verification stage for heuristic and ANN recognizers are 5.1% and 2.5%, respectively, i.e., the range for heuristic recognizers is more than double of the range for ANN recognizers. This implies that ANN recognizers have better generalization. In other words, ANN recognizers will have more consistent recognition performance than heuristic recognizers. However, the feature-based heuristic recognizer has the distinct advantage that the practitioners can clearly understand how a particular pattern is identified by the use of relevant shape features, which increases chances of its successful implementation. The best heuristic recognizer in terms of generalization is the recognizer No. 1.3, and the corresponding heuristic rules in the form of classification tree are shown in Fig. 9. It may be noted that the recognition performances of the proposed approach as presented in this paper are not truly comparable with the approaches of Pham and Wani (1997) or Hassan et al. (2003) because this approach is capable of recognizing eight types of patterns whereas the previous approaches can recognize six types of patterns only.

### 7.1. Confusion matrix

The confusion matrix is a table summarizing the tendency of the recognizer to classify a recognized pattern into a correct class or into any of the other seven possible (wrong) classes. Confusion matrices, as given in Tables 5 and 6, provide the overall mean percentages for confusions among the pattern classes for the six feature-based heuristic and six ANN recognizers, respectively. In other words, they are the mean of scores from 36 such matrices (6 recognizers $\times$ 6 testing sets).

Tables 5 and 6 show that there is confusion in the classification process for both types of recognizers. For both types of recognizers, there is a tendency for the shift patterns to be mostly confused with trend, and trend patterns with shift. This indicates that a new feature that is more powerful in discriminating shift and trend patterns, if can be identified, will be very useful. Downward shift patterns are the hardest to be classified for the heuristic recognizers (88.49%) as well as ANN recognizers (91.07%). The heuristic recognizers misclassify about 9.5% of downward shift patterns as downward trend patterns, whereas the ANN recognizers misclassify about 7.7% of downward shift patterns as downward trend patterns.

Generally, the results for classification of normal patterns in Table 5 (93.53%) and Table 6 (94.87%) suggest that the type I error performance for both types of recognizers does not seem to be very good. This is possibly due to the unpredictable structure of random data streams that make them relatively more difficult to be recognized compared to other unnatural patterns.

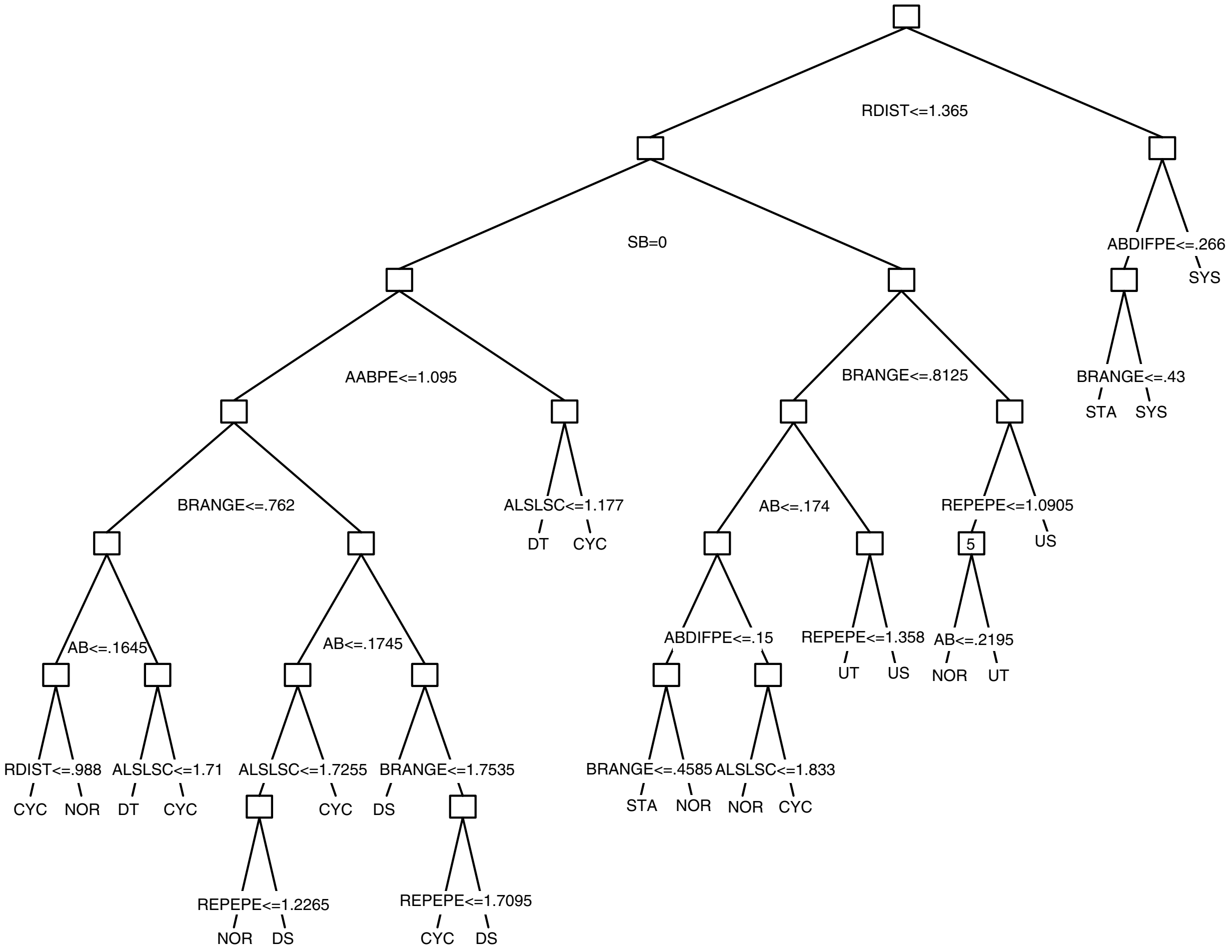


Fig. 9. Classification tree for recognition of control chart patterns.

Table 5
Mean percentage for confusion in heuristic recognizers

| True pattern class | Pattern class identified by heuristic recognizers | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | NOR | STA | SYS | CYC | UT | US | DT | DS |
| NOR | 93.53 | 2.73 | 0.69 | 0.44 | 0.76 | 0.24 | 1.16 | 0.44 |
| STA | 1.27 | 97.91 | 0.78 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |
| SYS | 0.20 | 1.56 | 98.24 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| CYC | 2.31 | 0.02 | 0.00 | 94.78 | 0.00 | 0.00 | 0.69 | 2.20 |
| UT | 0.98 | 0.00 | 0.00 | 0.00 | 98.40 | 0.62 | 0.00 | 0.00 |
| US | 0.22 | 0.00 | 0.00 | 0.00 | 7.24 | 92.53 | 0.00 | 0.00 |
| DT | 0.64 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 | 97.47 | 1.78 |
| DS | 0.07 | 0.02 | 0.00 | 1.91 | 0.00 | 0.00 | 9.51 | 88.49 |

## 8. Conclusion

A set of eight new shape features is presented that is useful for recognition of control chart patterns. Extraction of these features does not call for utilizing the experience and skill of the users and thus the CCP recognizer developed based of these features will be truly automated. The relative performances of the heuristic recognizers and ANN recognizers developed based on these features are extensively studied. In this study,

Table 6
Mean percentage for confusion in using ANN recognizers

| True pattern class | Pattern class identified by ANN recognizers | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | NOR | STA | SYS | CYC | UT | US | DT | DS |
| NOR | 94.87 | 1.40 | 0.67 | 0.00 | 0.87 | 0.13 | 1.93 | 0.13 |
| STA | 0.13 | 99.87 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| SYS | 0.47 | 0.07 | 99.40 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 |
| CYC | 0.47 | 0.00 | 0.00 | 98.73 | 0.00 | 0.00 | 0.27 | 0.53 |
| UT | 1.13 | 0.07 | 0.20 | 0.00 | 96.53 | 2.07 | 0.00 | 0.00 |
| US | 0.33 | 0.00 | 0.27 | 0.13 | 5.87 | 93.40 | 0.00 | 0.00 |
| DT | 1.40 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 96.60 | 2.00 |
| DS | 0.20 | 0.00 | 0.00 | 1.07 | 0.00 | 0.00 | 7.67 | 91.07 |

the feature-based ANN recognizers have achieved better recognition performance than the feature-based heuristic recognizers. The results also indicate that the feature-based ANN recognizers give more consistent recognition performance than the heuristic recognizers. However, the feature-based heuristic recognizer has the distinct advantage that the practitioners can clearly understand how a particular pattern is identified with the help of relevant shape features, which increases chances of its successful implementation. For both types of recognizers, there is a tendency for the shift patterns to be mostly confused with trend, and trend patterns with shift. Identification of a new feature that will be helpful in discriminating shift and trend patterns may be quite useful. One limitation of the proposed approach is that the CCP recognizer will be tailor-made for a given process since the magnitudes of all the features are not independent of process standard deviation.

It is important to note that the feature-based approaches for CCP recognition as available in the literatures are only focused on the recognition of basic six CCPs and moreover, require sufficiently larger number of observations for extraction of the related features. In this paper, the features are extracted from considerably smaller number of observations and the developed recognizers can recognize all the eight most commonly observed CCPs including stratification and systematic patterns.

## Appendix A. Generation of training and verification patterns

The following equations are used to generate different patterns for the training and testing data sets:

(a) Normal patterns

$$y_i = \mu + r_i\sigma \tag{9}$$

(b) Stratification patterns

$$y_i = \mu + r_i\sigma' \tag{10}$$

(c) Systematic patterns

$$y_i = \mu + r_i\sigma + d \times (-1)^i \tag{11}$$

(d) Increasing or decreasing trend

$$y_i = \mu + r_i\sigma \pm \mathrm{i}g \tag{12}$$

(e) Upward or downward shift

$$y_i = \mu + r_i\sigma \pm ks \tag{13}$$

(f) Cyclic patterns

$$y_i = \mu + r_i\sigma + a\sin(2\pi\mathrm{i}/T) \tag{14}$$

where $i$ is the discrete time point at which the pattern is sampled ($i = 1, \ldots, 32$), $k$ is 1 if $i \geqslant P$ (point of shift); otherwise $k = 0$, $r_i$ is the random value of a standard normal variate at $i$th time point and $y_i$ is the sample value at $i$th time point.

## References


Al-Assaf, Y. (2004). Recognition of control chart patterns using multi-resolution wavelets analysis and neural networks. *Computers and Industrial Engineering, 47*, 17–29.

Assaleh, K., & Al-Assaf, Y. (2005). Features extraction and analysis for classifying causable patterns in control charts. *Computers and Industrial Engineering, 49*, 168–181.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks/Cole Advanced Books and Software.

Cheng, C. S. (1997). A neural network approach for the analysis of control chart patterns. *International Journal of Production Research, 35*, 667–697.

Evans, J. R., & Lindsay, W. M. (1988). A framework for expert system development in statistical quality control. *Computers and Industrial Engineering, 14*(3), 335–343.

Grant, G. E., & Leavenworth, R. S. (1996). *Statistical quality control* (7th ed.). New York: McGraw-Hill.

Guh, R. S., & Tannock, J. D. T. (1999). Recognition of control chart concurrent patterns using a neural network approach. *International Journal of Production Research, 37*(8), 1743–1765.

Guh, R. S., Zorriassatine, F., Tannock, J. D. T., & O'Brien, C. (1999a). On-line control chart pattern detection and discrimination – a neural network approach. *Artificial Intelligence in Engineering, 13*, 413–425.

Guh, R. S., Tannock, J. D. T., & O'Brien, C. (1999b). IntelliSPC: a hybrid intelligent tool for on-line economical statistical process control. *Expert Systems with Applications, 17*, 195–212.

Guh, R. S., & Shiue, Y. R. (2005). On-line identification of control chart patterns using self-organizing approaches. *International Journal of Production Research, 43*, 1225–1254.

Hwarng, H. B., & Hubele, N. F. (1993). Back-propagation pattern recognisers for $\overline{X}$ control charts: methodology and performance. *Computers and Industrial Engineering, 24*(2), 219–235.

Haykin, S. (1999). *Neural network: a comprehensive foundation* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hassan, A., Nabi Baksh, M. S., Shaharoun, A. M., & Jamaluddin, H. (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research, 41*(7), 1587–1603.

Loh, W. Y., & Vanichestakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with discussion). *Journal of the American Statistical Association, 83*, 715–728.

Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815–840.

Morgan, J. N., & Messenger, R. C. (1973). *THAID: A sequential analysis program for the analysis of nominal scale dependent variables*, Technical report, Institute of Social Research, University of Michigan, Ann Arbor.

Montgomery, D. C. (2000). *Introduction to statistical quality control* (2nd ed.). New York: Wiley.

Nelson, L. S. (1984). The Shewhart control chart – test for special causes. *Journal of Quality Technology, 16*(4), 237–239.

Nelson, L. S. (1985). Interpreting Shewhart $\overline{X}$ control chart. *Journal of Quality Technology, 17*(2), 114–117.

Pham, D. T., & Oztemel, E. (1992a). XPC: an on-line expert system for statistical process control. *International Journal of Production Research, 30*(12), 2857–2872.

Pham, D. T., & Oztemel, E. (1992b). Control chart pattern recognition using neural networks. *Journal of System Engineering, 2*, 256–262.

Pham, D. T., & Wani, M. A. (1997). Feature-based control chart pattern recognition. *International Journal of Production Research, 35*(7), 1875–1890.

Pham, D. T., & Sagiroglu, S. (2001). Training multilayered perceptrons for pattern recognition: a comparative study of four training algorithms. *International Journal of Machine Tools and Manufacture, 41*, 419–430.

Pham, D. T., & Oztemel, E. (1994). Control chart pattern recognition using learning vector quantization networks. *International Journal of Production Research, 32*, 721–729.

Perry, M. B., Spoerre, J. K., & Velasco, T. (2001). Control chart pattern recognition using back propagation artificial neural networks. *International Journal of Production Research, 39*, 3399–3418.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Norstrand Reinhold.

Swift, J. A., & Mize, J. H. (1995). Out-of-control pattern recognition and analysis for quality control charts using LISP-based systems. *Computers and Industrial Engineering, 28*(1), 81–91.

Velasco, T., & Rowe, M. R. (1993). Back propagation artificial neural networks for the analysis of quality control charts. *Computers and Industrial Engineering, 25*(1–4), 397–400.

Western Electric. (1958). *Statistical quality control handbook*. Chicago, IL: American Telephone and Telegraph Company.

Wang, J., Kochhar, A. K., & Hannam, R. G. (1998). Pattern recognition for statistical process control charts. *International Journal of Advanced Manufacturing Technology, 14*, 99–109.

Yang, M. S., & Yang, J. H. (2002). A fuzzy-soft learning vector quantization for control chart pattern recognition. *International Journal of Production Research, 40*(12), 2721–2731.

Yang, J. H., & Yang, M. S. (2005). A control chart pattern recognition system using a statistical correlation coefficient method. *Computers and Industrial Engineering, 48*, 205–221.