Abnormal Patterns Detection in Control Charts using Classification Techniques

Zalinda Othman and Huda Fathi Eshames

School of Information Technology, Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia.
E-mail: zalinda@ftsm.ukm.my

Abstract

Any abnormal patterns show in Statistical Process Control charts imply the presence of possible assignable causes and variances that may lead to the process performance deterioration. Therefore, timely detection and recognizer of patterns in control charts are very important in the SPC implementation. This paper presents the performance of five classification methods on a set of large data for anomaly patterns detection in control charts. The control chart dataset has its specific features that need specific data preprocessing procedures. It is crucial and involves a number of stages of data preparation procedures. Firstly, the Principle Component Analysis is employed for similarity measure. Secondly, the Piecewise Aggregate Approximation and Symbolic Aggregate Approximation are used as data representation. The preprocessed data are fitted to the classification algorithms to extract important knowledge. The algorithms are support vector machine, decision tree, MLP networks, RIDOR algorithm and JRip algorithm. Numerical results showed that the JRip algorithm has the best performance compared to the others. It achieved highest detection accuracy about 99.66 % and the lowest error rate is 2.987.

Keywords: Control Chart Patterns Recognition, Statistical Process Control (SPC), Time Series Data Mining.

1. Introduction

Control charts are useful statistical tools for controlling and monitoring the quality characteristics of industrial processes. They are useful in determining whether a process is behaving as intended or if there are some unnatural causes of variation. Shewhart developed a control chart in 1931 [1]. Control chart provides a graphical representation of the process behavior that is used to monitor the system’s condition and for identifying instability conditions. The advantages of control charts can be further increased by adopting a control strategy that is, first, based on detecting patterns behavior (e.g. cyclic, trend, systematic changes) that can appear on the chart and second, lends itself to adaptive automation [2]. The main problem of most multivariate control charts is that they can detect an unusual event but do not directly provide the information required by a practitioner to determine which one of the variables or which subset of them causes the out-of-control signal. This raises the issue of identification in relation to multivariate control procedures. During diagnosis of the process behavior, the variations that may appear on the chart can be classified to various categories (e.g. patterns) based on the sources that cause them. Accurate classification of these patterns is vital for a production system to effectively monitor and reduce process variations. Chance/random variations are due to common causes (natural cause), and other variations are due to special causes (assignable cause). Common cause occurs naturally in the manufacturing process such as poor product design, inappropriate material, poor instruction and old machines. Western Electric [3] described in the handbook of natural process behaviour that most data points appear on the control charts near the centre line and have no points or actually very few (not exceeding control limits). Unnatural behaviour shows the points on a chart fluctuating widely (or not widely enough), or fails equilibrium about the centre line of the chart. Most commonly encountered patterns are shift, trend, normal, cycle, stratification, systematic and mixture that may appear on the control chart as defined in the Handbook by Western [3].

Control charts data are the time series data due to the large amount of data with the time stamp. The computing time is very complex. In order to reduce this computation time, the time series data need to be represented in high dimensional order. Most of the time series data are real valued data that require the data to be discredited [4]. A time series is a series of observations, $x_i(t): [i = 1, ..., n; t = 1, ..., m]$
made sequentially through time where, the measurements indexes made at each time point, \( t \). The resulting time series is univariate or multivariate respectively. It is called a univariate time series when \( n \) is equal to 1, and a multivariate time series (MTS) when \( n \) is equal to, or greater than 2. Control charts dataset that involves time is the MTS dataset [5]. The process behavior that appears on the charts indicates whether the quality of the manufacturing process is in control or not. Therefore, a good control chart monitoring should be able to detect the out of control situations and know when to stop the process.

This paper is organized as follows: the next section reviews the related research done on anomaly detection. Section 3 presents the methodology of our approach. Section 4 discusses the experiments performed and results obtained and finally, the last section concludes this study.

2. Related works

Anomaly detection can detect unknown variations, under a basic assumption that variations deviate from normal behavior. Patterns recognition can be defined as a classification of input data into identifiable pattern classes through the extraction of representative features of the data. There are several reported research papers on summarization and categorization of all possible patterns that are commonly appeared in SPC charts.

Over the years, rapid developments in computer technology motivate researchers to explore new methods based on CCP recognizer such as artificial neural network (ANN). Example work shown in Guh [6] that have used neural network (NN) and genetic algorithm (GA), where the GA was proposed for optimizing the NN recognizer. Later he [7] suggested ANN-based model for recognition of both mean and variance SSPs. He employed direct data and selected statistical features as input vector to ANNs. Another recent work that using ANN is by Ebrahimzadeh & Ranaee [8]. They designed an accurate, automatic system for recognition of CCPs that hybridized RBF wavelet methods and PSO approach. Literature reviews show NNs have shown capability as a suitable tool for detecting unnatural patterns in CCP. It is because the learning capability, speed and ease of implementation and can be used to deal with incomplete or noisy data.

Support vector machine (SVM) also has attracted researchers in CCPs recognition ([9] and [10]). Wang [11] proposed SVM to detect unnatural patterns in control charts such as, cyclic, shift and trend patterns. In the research work reported by Das & Banerjee [12], SVM with self-organized maps was used to build a CCP detector for online implementation. Zhiqiang and Yizhong [13] and Cheng et al. [14] suggested a model named least squares SVM to detect behaviour of variance shifts in bivariate processes.

Decision tree (DT) is well-recognised approach for its capability in detecting known attacks [15]. In CCPs pattern detection problem, several researchers have used DT. For example, Guh [16] proposed a hybrid learning-based model for recognition of CCPs. Guh's work used NN and DT to detect and identify eight types of control chart patterns. Guh & Shiue [17] developed an automatic system for CCP identification based on DT learning and back propagation networks (BPNs). Wang et al. [18] proposed independent component analysis with DT to identify various types of CCPs. Six types of concurrent control chart patterns were used to validate the proposed approach.

Different from the above approaches, this paper presents the time series preprocessing approaches (Principal Component Analysis (PCA) and Symbolic Aggregate Approximation (SAX) with Piecewise Aggregate Approximation (PAA)) to represent control charts data prior to detecting for anomaly patterns by using a number of approaches such as SVM, MLP networks, DT, Ridor algorithm and JRip algorithm.

3. Methodology

In this section, the control chart data preprocessing stages are discussed. It involves two main steps: the measure of similarity using the Principal Component Analysis (PCA) method and the transformation of these data using Symbolic Aggregate Approximation (SAX) with Piecewise Aggregate Approximation (PAA).
3.1 Similarity measure

The first step is to compute the similarity by using PCA. PCA is one of the mathematical techniques commonly used in the data mining community and the statistics for analysis of time series data. It is described by Dash et al. [19] as:

“It is an unsupervised linear feature reduction method, for projecting high dimensional data into a low dimensional space with minimum reconstruction error”.

This approach produces a reduced set of features that specify the maximal variances in the data as well as the discriminative features most adequate for classification, with minimum loss of information. In this section, the application of PCA in representing time series \( (T) \) of length \( n \) as a matrix \( X_{n \times p} \) will be explained, where \( n \) represents number of time instances (length of time series); whereas \( p \) is the number of time series data (number of variables being measured). The derivation of the new axes (components) is based on the computation of covariance matrix of the data as following equation (Tanaka et al. [20]):

\[
A = \begin{bmatrix}
\sum x_{1i}x_{1i} & \sum x_{1i}x_{2j} & \cdots & \sum x_{1i}x_{mi} \\
\sum x_{2i}x_{1i} & \sum x_{2i}x_{2j} & \cdots & \sum x_{2i}x_{mi} \\
\vdots & \vdots & \ddots & \vdots \\
\sum x_{mi}x_{1i} & \sum x_{mi}x_{2j} & \cdots & \sum x_{mi}x_{mi}
\end{bmatrix}
\] (1)

The calculation of the eigenvalues \( \lambda_i \) of this matrix is ordered as \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \). The \( M \) eigenvectors corresponding to the \( M \) largest eigenvalues of the covariance matrix is represented as \( e_1, e_2, \ldots, e_m \). Then the \( i^{th} \) principal component \( pc_{i,\lambda} \) is computed by using means of \( x_1, x_2, \ldots, x_n \).

\[
pc_{i,\lambda} = e_1\lambda_1(x_1 - \bar{x}) + e_2\lambda_2(x_2 - \bar{x}) + \cdots + e_m\lambda_m(x_m - \bar{x})
\] (2)

Then, the principal component is used to effectively transform the time series data from \( n \)-dimensional to \( m \)-dimensional space. Finally, the \( m \)-dimensional time series data \( (T) \) is illustrated as below:

\[
T = x_1, x_2, \ldots, x_n
\] (3)

\[
x_i = e_1\lambda_1(x_{1i} - \bar{x}) + e_2\lambda_2(x_{2i} - \bar{x}) + \cdots + e_m\lambda_m(x_{mi} - \bar{x})
\] (4)

PCA can detect the important coordinates that include characteristics of the original data \( T \), because the importance of each coordinate is represented in each coefficient \( e_i \). PCA can improve the performance of a detector as mentioned by Fengbin Zhang et al. [21] in their work on anomaly detection problem.

3.2 The combination of SAX and PAA

In this study, PAA with SAX is used for data representations. Firstly, the time series data are discretized by obtaining a PAA approximation. The idea of PAA representation is that it represents the time series data as a vector expression. This is obtained by dividing time series data into frames and calculating the mean value in each frame. It represents the time series data as a sequence of rectangle basis functions and then converts it into the discrete symbolic sequences, as refer in Eq. (5). In this equation, a time series data \( T = x_1, x_2, \ldots, x_n \) of length \( n \) can be represented as \( w \), i.e. the number of PAA segments/frames representing time series \( (m < n) \), and \( X_j \) is the average value of the \( j^{th} \) segment and a vector \( X_i \) is calculated as follows:
\[ X_i = \frac{1}{n} \sum_{j=n-i+1}^{n} X_j \]  \hspace{1cm} (5)

Secondly, after PAA represented of each time series data by vector of \( w \)-dimension, wherein, “breakpoints” can be determined by mapping the vector \( X \) into a sequence of SAX symbols. Lin et al. [4] define breakpoint as; “Breakpoints are a sorted list of numbers \( B = \beta_0, \beta_1, \ldots, \beta_{a-1} \) such that the area under a \( N(0,1) \) Gaussian curve from \( \beta_i \) to \( \beta_{i+1} = \int_{-\infty}^{\beta_{i+1}} \phi(x) dx \) (\( \beta_0 \) and \( \beta_{a} \) are defined as \( -\infty \) and \( \infty \), respectively)”. These breakpoints may be determined by looking them up in a statistical table as shown in Table 1. The data is normalized into fixed number of states according to the predetermined breakpoints. This produces an equal-sized area under Gaussian distribution curve to map PAA to SAX symbols. By referring to the SAX Gaussian Distribution Table, as in Table 1, the corresponding \( \beta \) and related values of \( a \) (from 3 to 10) for the breakpoints can be determined. Once the breakpoints have been obtained we can discretize a time series in the following manner. All coefficients of PAA that are below the smallest breakpoints are mapped to the symbol “a” and all PAA coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint are transformed to the symbol “b” and so on. This idea is shown as in Figure 1. It is noted in this figure that two breakpoints have been obtained, therefore 3 symbols a, b, and c are used. In this example, the time series of length \( n = 128 \) are discretized by first obtaining a PAA approximation for reduction to \( w = 8 \). Then, by using the predetermined breakpoints PAA coefficients are map into SAX symbols. In the figure, by using \( a = 3 \), the time series is map to the word baabccbc.

### Table 1. The SAX Gaussian Distributions

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_2 )</td>
<td>-0.43</td>
<td>-0.67</td>
<td>-0.84</td>
<td>-0.97</td>
<td>-1.07</td>
<td>-1.15</td>
<td>-1.22</td>
<td>-1.22</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.43</td>
<td>0</td>
<td>-0.25</td>
<td>-0.43</td>
<td>-0.57</td>
<td>-0.67</td>
<td>-0.76</td>
<td>-0.84</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>0.67</td>
<td>0.25</td>
<td>0</td>
<td>-0.18</td>
<td>-0.32</td>
<td>-0.43</td>
<td>-0.52</td>
<td></td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>0.84</td>
<td>0.43</td>
<td>0.18</td>
<td>0</td>
<td>-0.14</td>
<td>-0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_6 )</td>
<td>0.97</td>
<td>0.57</td>
<td>0.32</td>
<td>0.14</td>
<td>0</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_7 )</td>
<td>1.07</td>
<td>0.67</td>
<td>0.43</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_8 )</td>
<td>1.15</td>
<td>0.76</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_9 )</td>
<td>1.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_{10} )</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** PAA transformation to SAX symbols
4. Classifier

The proposed classification algorithms in this study are described in this section:

A. Decision Tree

Decision Tree (DT) is one of the popular and powerful learning techniques. It is due to its various characteristics that include simplicity, comprehensibility, less parameters, and its ability to handle mixed-type data. A DT is induced from a set of labeled training instances represented by a couple of attribute values and class labels. Its learning process is typically a greedy, top-down and recursive process starting with the entire training data and an empty tree. An attribute that best partitions the training data is chosen as the splitting attribute for the root. The training data are then partitioned into disjointed subsets satisfying the values of the splitting attributes [22].

B. Support Vector Machine

Support Vector Machine (SVM) is considered one of the successful learning algorithms proposed in the application of classification, regression and detection tasks. Referring to Chang & Chang [23], SVM is a powerful machine-learning tool. It has the capacity to represent non-linear relationships and producing models that generalize well to unseen data. The main goal of applying SVM for solving classification problems is to transform the input space to a higher dimensional feature space through a non-linear mapping function.

C. Multilayer Perceptron Neural Networks

Multilayer Perceptron (MLP) Neural Networks have been used over the years as a powerful learning algorithm for solving a wide variety of problems. Much progress has been made in improving performance and in understanding how these neural networks operate. MLP neural networks consist of units arranged in layers. Each layer is composed of nodes and in the fully connected networks. Each node connects to every node in subsequent layers. Each MLP is composed of a minimum of three layers consisting of an input layer, one or more hidden layer(s) and an output layer. The input layer distributes the inputs to subsequent layers. Input nodes have linear activation functions and no thresholds. Each hidden unit node and each output node have thresholds associated with them in addition to the weights [24].

D. JRip algorithm

Extended Repeated Incremental Pruning (JRip) approach was proposed by Cohen [25] as an optimized version of IREP. RIPPER extracts the rules directly from the data. It implements a propositional rule learner, “Repeated Incremental Pruning to Produce Error Reduction” (RIPPER). Before building a rule, the current set of training examples are partitioned into two subsets, a growing set (usually 2/3) and a pruning set (usually 1/3). RIPPER builds a rule set by repeatedly adding rules to an empty rule set until all positive examples are covered. Rules are formed by greedily adding condition to the antecedent of a rule (starting with empty antecendent) until no negative examples are covered. After growing a rule from the growing set, condition is deleted from the rule in order to improve the performance of the rule set on the pruning examples. To prune a rule, RIPPER considers only a final sequence of conditions from the rule, and selects the deletion that maximizes the function. As in Eq. (6):

\[ v(Rule, prpos, prNeg) = \frac{p-n}{p+n} \]  

Where Rule is the set of rules, prpos is the total number of examples in the considered cluster, prNeg is the total number of examples in the cluster not considered and p, n are the number of (prpos, prNeg) examples covered by Rule.

E. Ridor algorithm

Ridor algorithm is the implementation of a RIpple-DOwn Rule learner proposed by Gaines and Compton [26]. It generates a default rule first and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the “best” exceptions for each exception and iterates until pure. Thus it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default. IREP is used to generate the exceptions.
5. Simulation results

*Synthetic Control Chart Time Series* (SCCTS) datasets are taken from the UCI Machine Learning Repository website [27]. The datasets consist of 600 examples of control charts. These examples are divided into six different classes (patterns). Each class consists of 100 time series and the length of each time series is equal to 60 (60 numerical attributes). These classes are cyclic, normal, increasing/decreasing trend, upward/downward shift and generated according to the six equations as given in Pham and Oztemel [28] and Pham and Chan [29].

In this paper, the raw collected data are cleaned according to time series pre-processing procedures. Initially, PCA is applied to obtain reduced uncorrelated attributes by specifying maximal variances in the data with minimum loss of information, and it implies the process of data compress. This approach transformed the data from high-dimensional data to low-dimensional data. The results obtained shown that the PCA has achieved optimization efficiency in dimensions reduction. It is reduced number of dimensions/attributes dataset from 60 attributes to 40. The values of PCA are then being transformed to symbols by using PAA and SAX. Based on Gaussian distribution curve, five symbols are chosen. In this study, the symbols are represented in numbers. In this paper, four breakpoints are obtained. The control chart data are mapped to numbers using \( a = 5 \). The data are transformed into number 1 to 5. By using the above steps, a program is developed for data preprocessing by using Visual Basic for Applications (VBA). These procedures achieved the effective properties in pattern representation and reached for building a good classification model. The preprocessed data are mined using five classification techniques as discussed in previous section. The performance is measured by the anomaly detection accuracy and error average with time taken in building the model. The mining process uses 10 folds cross validation procedure with various allocations of training and testing set. In this study, the WEKA and ROSETTA are used as a mining and modeling tool.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Detection Accuracy %</th>
<th>Error Rate</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>99.44</td>
<td>3.305</td>
<td>0.04</td>
</tr>
<tr>
<td>SVM</td>
<td>96.66</td>
<td>9.708</td>
<td>1.84</td>
</tr>
<tr>
<td>JRip</td>
<td>99.66</td>
<td>2.987</td>
<td>0.12</td>
</tr>
<tr>
<td>Ridor</td>
<td>99.44</td>
<td>3.061</td>
<td>0.17</td>
</tr>
<tr>
<td>MLP</td>
<td>98.33</td>
<td>5.584</td>
<td>16.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Mean Absolute Error</th>
<th>Root Mean Squared Error</th>
<th>Relative Absolute Error (%)</th>
<th>Root Relative Squared Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.233</td>
<td>0.311</td>
<td>80.067</td>
<td>83.230</td>
</tr>
<tr>
<td>JRip</td>
<td>0.006</td>
<td>0.036</td>
<td>2.189</td>
<td>9.717</td>
</tr>
<tr>
<td>MLP</td>
<td>0.012</td>
<td>0.067</td>
<td>4.226</td>
<td>17.890</td>
</tr>
<tr>
<td>DT</td>
<td>0.004</td>
<td>0.043</td>
<td>1.528</td>
<td>11.645</td>
</tr>
<tr>
<td>Ridor</td>
<td>0.002</td>
<td>0.043</td>
<td>0.666</td>
<td>83.230</td>
</tr>
</tbody>
</table>
The experimental results indicated that the proposed techniques have produced the model with high accuracy and lower error rate with lower total time taken to build model. Based on Figure 2, 3, Table 2, and 3, it is noted that each proposed algorithms achieved high detection accuracy. The highest accuracy is 99.66%, produced by JRip and followed by DT and Ridor are about 99.44%, then MLP networks is 98.33% detection accuracy and SVM is about 96.66%. The time requires to build the model is also a parameter for comparing the classification techniques. It can be observed from Figure 3 that a DT has taken the shortest building time that is around 0.04 seconds compared to the others. JRip requires around 0.12 seconds, Ridor 0.17 seconds and SVM is 1.84 seconds for building the model. MLP networks take the longest time to build the model that is around 16.16 seconds. From Table 2, it is discovered that the highest error is found in SVM with an average errors of 9.708, while the rest of the techniques ranging from 2.987 to 5.584. The technique with lower error rate has a powerful classification capability. In addition, more potential in terms of control chart patterns recognizer. Besides the classification capability, the anomaly in CCPs can be detected by observing the generated rules from the classification model. Table 4 provides an overview of the rules generated from the selected JRip classification model. Rule induction algorithms generate a model as set of rules. The rules are in the standard form of If-Then rules. The number of rules obtained varies from one fold/model to another. The rules generated in the best model cover all instances and features. These rules might lead to a good decision and conclusion. These rules have an intuitive interpretation. For example, the first rules states: (T1>= 3) => class = 1 (50.0/0.0), if the T1 (the first attribute in the data, it represents a period of time) occurs is the larger than or is equal to 3 (a point representing quality characteristics of sample taken from the manufacturing process at a period of time, then class is equal 1 (type of pattern/class is normal). This means that process is in control. The number in parenthesis indicates how many on the training examples were covered successfully by this rule. For example, for the first rule, 50 examples were covered by this rule correctly (covers instances with total weights of 50.0), out of which there are instances with weights of 0.0 are misclassified.
Table 4. Rules generated from the best classification model based on JRip Algorithm

<table>
<thead>
<tr>
<th>Class/Pattern</th>
<th>Related rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>=&gt; Class = 1 (T1 &gt;= 3) =&gt; Class = 1 (50.0/0.0)</td>
</tr>
<tr>
<td>Cyclic</td>
<td>=&gt; Class = 2 (T4 &lt;= 2) =&gt; Class = 2 (47.0 / 1.0) (T2 &lt;= 1) =&gt; Class = 2 (4.0 / 0.0)</td>
</tr>
<tr>
<td>Increasing trend</td>
<td>=&gt; Class = 3 (56.0/3.0)</td>
</tr>
<tr>
<td>Decreasing trend</td>
<td>=&gt; Class = 4 (T1 &gt;= 5) =&gt; Class = 4 (48.0/0.0)</td>
</tr>
<tr>
<td>Upward Shift</td>
<td>=&gt; Class = 5 (T1 &lt;= 2) and (T7 &gt;= 5) =&gt; Class = 5 (31.0/0.0) (T5 &lt;= 1) and (T1 &lt;= 2) =&gt; Class = 5 (14.0/0.0)</td>
</tr>
<tr>
<td>Downward shift</td>
<td>=&gt; Class = 6 (T1 &gt;= 4) and (T7 &lt;= 2) =&gt; Class = 6 (39.0/1.0) (T5 &gt;= 5) and (T1 &gt;= 2) =&gt; Class = 6 (11.0/0.0)</td>
</tr>
</tbody>
</table>

6. Conclusions

As a conclusion, the experimental results show that data mining and machine learning can be extremely beneficial in generating quality knowledge and may benefit the manufacturers. This work has achieved its proposed goals in implementing suitable approaches to preprocess control chart data prior to the classification task. In addition, to investigate and evaluate five selected classification techniques. The performances of these techniques are measured based on recognition accuracy, error rate and the time taken to build model. The preprocessing tasks proposed have shown that with proper preparation of control chart data, the classification techniques are suitable for solving this problem. Through the results obtained, it can be clearly seen that all the values of recognition accuracy and average errors indicate that the best classification model is obtained by using JRip algorithm. It has proved effective and fast to use even with large and noisy datasets. Therefore, it gave the highest accuracy, and lowest error rate with shortest time to build the model.

Future work can be carried out in several areas, including using real time data, since it provides a genuine test for verifying the classification methods in SPC implementation. Studying the statistical characteristics of the different features of data may provide important guidelines developing and improving pattern recognition applications in SPC. In this study, only six types of abnormal patterns were studied. Therefore, the future work can be extended to add more patterns in the control scheme, such as stratification or mixture. In addition, several efforts have focused on the development of methodologies for the mining process of the control chart data in the last decade. However, relatively few methods have been developed to classify the data with an efficient way. Therefore, it is necessary to conduct further research on developing a classification model for the process control data.

7. Acknowledgment

The authors wish to thank UKM for its financial support under UKM-TT-07-FRGS0158-2010 grant.

8. References