Control Chart Pattern Recognition: A Time Series Data Mining Approach

Leonidas Karamitopoulos
University of Macedonia
156 Egnatia St.
GR-54006 Thessaloniki
lkaramit@uom.gr

Georgios Evangelidis
University of Macedonia
156 Egnatia St.
GR-54006 Thessaloniki
gevan@uom.gr

Dimitrios A. Dervos
ATEI of Thessaloniki
P.O. Box 141
GR-57400 Sindos
dad@it.teithe.gr

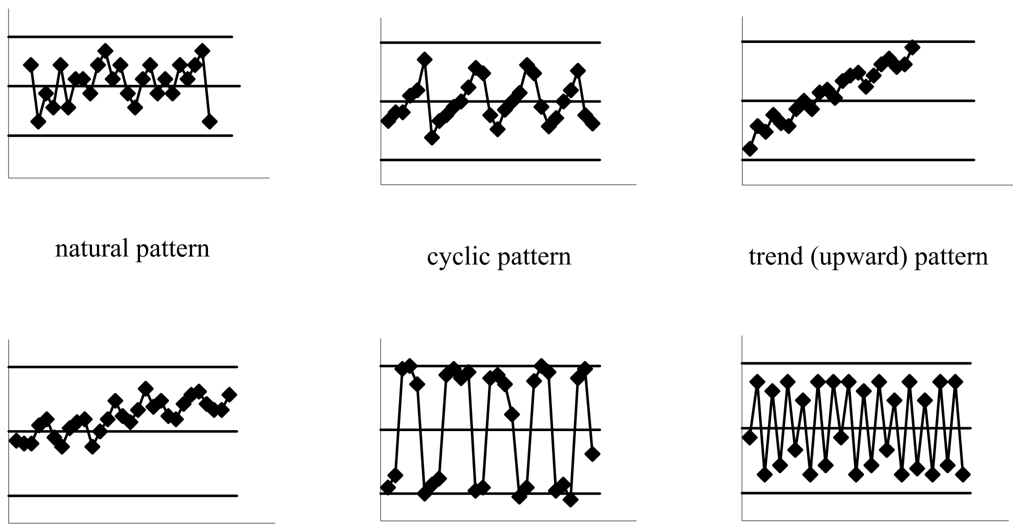
Abstract. The need for constantly monitoring and measuring the level of quality in manufacturing industries led to the development and application of specific procedures, one of which was the Statistical Process Control (SPC) that mainly involved the construction and interpretation of control charts. Although the traditional methods have been proven to be efficient, advances in technology of generating and storing huge amount of data emerged the need of developing new techniques and tools capable of intelligent process and analysis in real time. Expert Systems and Neural Networks have been investigated and tested for their usefulness in SPC and more specifically in control chart pattern recognition. The Data Mining field has also attracted a lot of attention during the last decade since it involves techniques that assist in analyzing enormous amount of data. In this paper, we investigate the possible contribution of Data Mining techniques in control chart pattern recognition. We discuss the general issues of a Data Mining approach within the SPC context namely, the choice of representation, similarity measures and algorithms. Experiments were conducted on synthetic datasets taking into consideration many parameters such as noise, magnitude of deviations and presence of in-control data. Two different representations were selected for their different characteristics: Feature-based and SAX. Decision Trees were selected as the classification method to be applied for its simplicity and ability to produce understandable rules. The evaluation of the various approaches is based upon the classification accuracy rates. Results show that although the selected models are very simple, they can adequately classify different types of control chart patterns. Although, the trade-off between model complexity and accuracy rate is not investigated in this paper, results indicate that Data Mining may contribute significantly in control chart pattern recognition.

Keywords. Control charts, time series, data mining, classification, pattern recognition.

1. INTRODUCTION

Statistical Process Control (SPC) consists the basic method of quality control not only in manufacturing industries but also in services. Contrary to the traditional methods of quality testing of the final product, SPC focuses in the whole process, from the raw materials to the final product. The main goals of SPC are: (a) monitoring (usually by sampling) the main parameters of the process, (b) detecting process deviation, and, (c) diagnosing and taking corrective action [15]. Although SPC was technically and theoretically developed in 1931 [21], it was adopted from the industries worldwide during the last two decades being one of the most effective tools in Total Quality Management (TQM) [4] [5].

Although there are several tools that should be utilized together, control chart is the most powerful tool in order to achieve SPC goals. The three fundamental uses of a control chart are: (a) reduction of process variability, (b) monitoring and surveillance of a process and (c) estimation of product or process parameters. Basically, a control chart is the graphical representation of a time series. An important (in quality context) variable of a process is being measured continuously for a specific time interval and the corresponding measurements are being plotted in a diagram over time. The unique characteristic of a control chart is that at the same diagram a lower control limit (LCL) and an upper control limit (UCL) are being drawn. The calculation of these limits is based on statistical theory. As long as the points (measurements) fall randomly within these limits the process is said to be in statistical control. The variation is due only to chance causes (common causes). When one or more points fall beyond the control limits or the plotted points exhibit some non-random pattern of behavior, the process is said to be out of control. In this case, variation is due also to the presence of assignable causes (special causes). Western Electric [23] determined fifteen different patterns that may be encountered in practice. Some of them are shown in Figure 1. The traditional methods of SPC have been proven to be very effective in detecting out of control conditions, especially when one or more points fall beyond the control limits. Problems arise when non-random (unnatural) patterns appear, since this involves the ability or experience of an operator to identify and classify them into the appropriate class of assignable causes. The SPC approach to this problem was the development of sensitizing rules (run rules or zone tests) in order to assist operators in identifying these patterns [18] [19]. However, the more rules are applied to a chart, the more complicated the decision process becomes and the number of false alarms may be increased resulting in a less effective SPC procedure.



shift (upward) pattern   mixture pattern   systematic pattern

Figure 1. Examples of Control Chart patterns

Nevertheless, new technologies of collecting and storing data produce huge amount of data, emerging the need for a new generation of techniques and tools capable of intelligent process and analysis in real time.

Artificial intelligence (AI) techniques have been widely investigated in order to assist in control chart pattern recognition (CCPR) and further automate the process control procedure. In particular, Expert Systems have been developed for the purpose of automating the interpretation and diagnosis of control chart information. The main advantages are the ease of development and transparent reasoning. Several Expert Systems have been proposed [6] [10] [22] that primarily perform the tasks of selecting, constructing and interpreting an appropriate control chart, and diagnosing the possible assignable causes (if present). For interpretation and diagnosis, sensitizing rules are incorporated to these systems resulting into the problem of false alarms (misclassification) as stated above. Neural Networks (NN) have also been investigated within SPC context, mainly for the purpose of identifying unnatural patterns in control charts. A general overview is provided in next section.

Finally, in the last decade, the Data Mining field has also attracted a lot of attention since it involves techniques that assist in analyzing enormous amount of data [7]. Time Series Data Mining (TSDM) is comprised by special purpose techniques that take into consideration the temporal nature of the data. Although one of the tasks of TSDM is classification and the control charts are constructed by time series, there is not much research in the possible contribution of TSDM to control chart pattern recognition.

The rest of this paper is divided into four further sections. In Section 2 we provide a general overview of the AI literature related to Expert Systems and Neural Networks. Section 3 describes the basic components of Time Series Data Mining methods. In Section 4, we present a TSDM approach in control chart pattern recognition. Section 5 describes the experiments that were performed and discusses the corresponding results. Finally, a conclusion is presented in Section 6.

1.1.1 NEURAL NETWORKS AND SPC PATTERN RECOGNITION

Neural networks have been applied to many applications in diverse areas such as science, business, manufacturing and biology. The main tasks that NNs can accomplish are classification (supervised and unsupervised) and prediction. These models mimic the operation of the neural connections in human brain by learning from data and developing an ability to generalize, similarly to the way humans learn from experience. An artificial neural network consists of a number of interconnected processing units called neurons, since the basic modeling is based on biological neurons. Neurons are organized into layers: input, hidden (one or more layers) and output. Each unit receives a number of input signals and combines them into a single output value by using an activation function. Each input signal is assigned a relative weight, and the weighed sum of these input signals constitute the effective input to the processing unit. From this result, an output value is calculated and is transferred to the output of the unit. Depending on the interconnection architecture, the output signal may be sent to other processing units as input signal. The network starts learning (adjusting the weights) by receiving training data with known outputs and by following a specific learning strategy.

Since the problem of control chart pattern recognition can be assumed as a classification problem, NN technology has been investigated for its usefulness in improving automation in SPC systems. There has been a lot of research in control chart pattern recognition using neural networks [2], [11], [12]. Different models have been implemented and tested in the context of pattern recognition in SPC. The procedure of the construction of these models consists of common steps and decisions to be made, namely, the representation of input data, the selection of the architecture and topology of the network, the choice of the activation function, the learning and training strategies. These decisions often interrelate to each other.

Within control chart pattern recognition context, there are many diverse approaches in representation of input data such as using raw data [20] or features extracting from data [17]. Several transformations also may be performed such as standardization, zoning, scaling and using continuous (as opposed to binary) representation [24]. The most frequently selected model is the Multi-Layer Perceptron feed forward fully connected network along with the back propagation rule and the sigmoid function. There are several training and testing strategies with respect to the window size under investigation, the size of the training dataset and the presentation order of the training patterns. The evaluation of the performance is mainly assessed by the classification accuracy, the average run length (ARL), and, the Type I and II errors. An analytical review of Neural Networks in control chart pattern recognition can be found in [24]

2. TIME SERIES DATA MINING

The Data Mining (DM) field involves techniques and algorithms capable of efficiently extracting patterns from large databases that can potentially constitute knowledge. The primary goals of DM methods are the description of a particular dataset (often huge) and the prediction of future values of interest based on already known values from a database. According to Fayyad et al [7], the basic data mining tasks are: classification, regression, clustering, summarization, dependency modeling and change and deviation detection. Time Series Data Mining (TSDM) is a relatively new field that is comprised by DM methods adjusted in a way that they take into consideration the temporal nature of data. According to the research in this field, the main tasks of TSDM methods are: indexing, clustering, classification, novelty detection, motif discovery and rule discovery [13].

Although some of these tasks are similar to the corresponding DM tasks, the temporal aspect arises some special issues to be considered and/or imposes some restrictions in the corresponding applications. First, in most of the above tasks, apparently, it is necessary to define a similarity measure between two time series. This issue is very important in TSDM since it involves a degree of subjectivity that might affect the final result. A lot of research was focused on defining different similarity measures in order to improve the performance of the corresponding methods. Although the most well-known measure is the Euclidean distance, many other measures have been proposed in the literature such as Dynamic Time Warping, Longest Common Subsequence measures with local scaling, and Longest Common Subsequence measures with global scaling to name a few. A thorough presentation and literature review of time series similarity measures can be found in [9].

A second issue that arises in TSDM and interrelates to the selection of a similarity measure is the representation of a time series. Since the amount of data may range from a few megabytes to terabytes, an appropriate representation of the time series is necessary in order to manipulate and analyze it efficiently. One objective is to reduce dimensionality which is very high in time series and thus deal with the problem of the "dimensionality curse" that appears frequently within real world DM applications. Another objective is to take advantage of the specific characteristics of a representation that make specific methods applicable. As with similarity measures, there are a lot of proposed representations in the literature, including the Discrete Fourier Transform (DFT), the Discrete Wavelet Transform (DWT), the Piecewise Aggregate Approximation (PAA), the Adaptive Piecewise Constant Approximation (APCA) and the Singular Value Decomposition (SVD) [14]. Finally, there are several symbolic representations based on many different discretization techniques [3].

3. A PROPOSED TSDM APPROACH

The case of control chart pattern recognition can be considered as a time series classification problem. As it was mentioned before, one possibility of a process being out of control is when the plotted points exhibit some non-random pattern of behavior, specifically, one of the fifteen patterns determined in [23]. Thus, given a set of pre-defined classes (possible patterns), the goal is to assign a time series into the most appropriate class.

A TSDM approach should involve the selection of (a) a representation scheme, (b) a similarity measure, and, (c) a data mining method for classification. These three decisions interrelate to each other and, thus, they should be considered concurrently. There are many proposed representation schemes and similarity measures in the literature, as it was stated in the previous section. Moreover, besides Neural Networks, there are many classification procedures available such as tree models, linear discriminants, logistic discriminant analysis, nearest neighbor methods and the naïve Bayes model, to name a few. Although one of the main tasks of TSDM is classification, there is not much research in the possible contribution of TSDM to control chart pattern recognition. It is not our intension to investigate and compare all available classification procedures in this paper, but to indicate the potential usefulness of TSDM methods in SPC.

4.1 Representation schemes

Three cases are considered in this paper: avoid the application of any representation scheme, meaning that we implemented the proposed method on raw data, and apply two representation schemes, namely, SAX and Feature-Based (FB). Since we intended to apply a classification method on relatively short time series (length of 56), the first choice was adequately applicable and was made for comparison reasons. The other two choices were made based on the fact that they differ significantly in their approach. The first one is a symbolic representation that transforms the original data into a vector of discrete symbols where the transformed values are of ordinal level retaining the time aspect of the original data. Also, this is a relatively recent proposed representation scheme in the TSDM community [14]. The second one is based on the calculation of statistical measures and transforms the original data into a vector of real values where the transformed values do not retain the time aspect of the original data.

Lin et al. proposed a representation of time series, called Symbolic Aggregate approXimation (SAX) that achieves dimensionality reduction and handles streaming data efficiently. The symbolic nature of the representation allows the application of methods and data structures appropriate for discrete data, such as Decision Trees. Last but not least, a similarity measure can be defined that lower bounds a distance measure defined on the original time series. This is an important property within Data Mining context since it allows dealing efficiently with huge amount of data [14]. The procedure that SAX method proposes is as follows. First, the time series are normalized to have a mean of zero and a standard deviation of one. Second, the time series are transformed further by applying a Piecewise Aggregate Approximation (PAA). The PAA technique divides a time series of length m into k consecutive windows of equal-width and calculates the corresponding mean for each one. The series of these means is the new representation of the original data. The user must define the value of the parameter k. Third, taking advantage of the fact that the transformed series follows the normal probability distribution, each element is mapped to a symbol using the properties of this distribution. The user must assign a value, in a second parameter α, which defines the alphabet size (number of symbols to be used). Then, the area under the normal curve is divided into α areas of equal size (meaning that the corresponding probabilities will be equal for each symbol) and each one of them is assigned to a symbol. Finally, an element of the series, which falls into an interval that corresponds to a specific area, is mapped to the area's symbol.

Nanopoulos et al. proposed the use of statistical features for time series classification, based on their employment in image processing. In their work [17], the original time series x_t of length n was transformed to a series y_t of length n-1 by using a user-defined value D and the following equation:

$$y_t = x_{t+D} - x_t, \qquad 1 \leq t \leq n - D \qquad (1)$$

As Nanopoulos et al. state, this series (y_t), contains information about the shape of the original time series and also acts as a noise filter. The selected statistical features are the mean value, standard deviation, skewness and kurtosis for the original (x_t) and the transformed (y_t) time series. These are statistical measures, which attempt to describe a dataset with respect to the central tendency and dispersion of its values along with the shape of their distribution. Consequently, a time series is transformed to a vector of eight real values (statistical features of the original and the transformed series).

4.2 Similarity measures

The similarity measure to be chosen is interrelated to both representation scheme and the classification method. In this paper, we have chosen a classification method that is decision tree-based and not distance-based. Consequently, there is no consideration on the similarity measure to select. Nevertheless, in the case of applying a distance-based classification method, the SAX representation can be applied with a similarity measure that lower bounds a distance measure defined on the original time series, as it was mentioned in the previous section.

4.3 A DM method for classification

In this paper, we consider a specific example of a TSDM approach in control chart pattern recognition, chosen for its simplicity and its ability to generate understandable rules. By far, the most simple classification method is Decision Trees (DT) [16]. The generation of a DT is developed in two phases: tree construction and tree pruning. During the first phase, starting from the root node and proceeding in a top-down manner, the training examples are partitioned recursively based on selected attributes. An internal node is mapped to an attribute and denotes a test on this attribute, a branch is mapped to an outcome of the test and a leaf node is mapped to a class. During the second phase, the tree is pruned by removing branches that reflect noise or by removing subtrees in order to avoid redundant comparisons.

The main advantages of DTs are their ability to produce understandable rules, their low computation requirements, and their ability to handle both continuous and categorical variables, although in the first case transformations are usually needed. Moreover, there is no need to determine a similarity measure and this results to a simpler procedure. There are some disadvantages though. DTs may be quite large and since they are constructed from training data, overfitting is possible. Tree pruning usually overcomes these problems. The decision tree approach is to divide the search space into rectangular regions and classify an object based on the region into which it falls. This implies that it is not an appropriate approach for all classification problems and also that it ignores correlation among attributes.

4. EXPERIMENTAL RESULTS

The aim of this paper is to investigate the possibility that TSDM techniques can contribute to control chart pattern recognition rather than apply and compare all potential approaches from this field. In order to conduct experiments, six control chart patterns (the most common types encountered in practice) were selected: natural, cyclic, upward trend, downward trend, upward shift and downward shift. Two datasets were generated each consisting of 1800 time series of length 56 (300 from each one of the six patterns), by using the following equations:

$$x_t = \mu + r_t \qquad \text{(natural)} \qquad (2)$$

$$x_t = \mu + r_t + A \sin(2\pi t/\Omega t) \qquad \text{(cyclic)} \qquad (3)$$

$$x_t = \mu + r_t \pm bt \qquad \text{(trend)} \qquad (4)$$

$$x_t = \mu + r_t \pm ps \qquad \text{(shift)} \qquad (5)$$

where,

x_t: time series value

μ: process mean ($\mu = 30$ in this paper)

σ: standard deviation when process is in-control ($\sigma = 2$ in this paper)

r_t: common cause variation, following a normal distribution with mean of zero and standard deviation in terms of σ (0.1σ, 0.3σ and 0.5σ)

A: cyclic amplitude in terms of σ (1σ ,2σ and 3σ)

Ω: period of cycle (Ω = 8 in this paper)

b: trend slope in terms of σ (0.1σ , 0.2σ and 0.3σ)

p: parameter that determines the shift position (p = 0 before shifting, p = 1 after shifting) while the shift position is set to be uniformly distributed between the 21st and 40th time points

s: shift magnitude in terms of σ (1σ , 2σ and 3σ)

The difference between the two datasets is in the presence of in-control data. In the first dataset, where in-control data is not included, the disturbance starts at the first time point for the cyclic and trend patterns, whereas in the second dataset, where in-control data is included, cyclic and trend patterns are generated by setting the starting time that the disturbance occurred uniformly distributed between the 21st and 40th time points. There is no clear-cut answer as to the most appropriate number of in-control data and the starting position of a disturbance [1]. The choice of the time series length is application-dependent. The process itself and the monitoring measurements affect the sampling frequency and, thus, the window size to be selected. In this paper, we considered time series of individual measurements and the length was determined to be 56. One reason is that it is a rational choice for applications where the frequency of sampling is high (e.g., every one second). Another reason is that it facilitates the specific experiments where we have included in-control data. The above parameters were selected to ensure that the majority of the generated patterns fluctuated within ±3σ, since the identification of a deviation beyond these limits can be easily identified as out of control incidence [8].

In the previous section, two different representation schemes were selected for the classification task. The SAX scheme has two parameters: the length of the transformed series (k) and the alphabet size (α). In our experiments k was set equal to 8 in order to achieve the same dimensionality reduction with the feature-based (FB) representation scheme, and α was set equal to 6, 8 and 10 resulting into three different representations. The FB scheme has one parameter D, which was set equal to one. Moreover, Decision Trees have been chosen as the Data Mining classification method to be applied. Three of the most popular algorithms have been applied in both datasets: CHAID (Chi-squared Automatic Interaction Detector), CRT (a version of the Classification and Regression Trees algorithm) and QUEST. The aim of these experiments was to compare the performance of various classification algorithms using different representations rather than determine the optimal performance. SPSS 13.0 was utilized in order to perform these experiments and its defaults were selected. Specifically, the maximum number of levels in the tree below the root node was set to 3 for CHAID and 5 for CRT and QUEST, while the minimum number of examples was set to 100 in the parent node and 50 in the child node. The corresponding trees were pruned with the maximum difference in risk to be of one standard error. These growth limits criteria and the pruning procedure aim at constructing simpler trees, consequently simpler rules, and dealing with the problem of overfitting. The evaluation measure is the classification accuracy for the testing dataset. The validation is based on randomly splitting (approximately in half) the dataset into training and testing samples. Each algorithm was repeated ten times and the corresponding values of mean and standard deviation were calculated for classification accuracy.

The results of the experiments are presented at Table 1. One important observation is that the classification rates are (in all cases except CRT with raw data and CRT with feature-based representation) lower than 90%, while the corresponding rates in the literature are above 90%. This is due to the strict criteria we used for the growth of the tree, as it was described earlier. These experiments resulted in fairly simple trees since the number of nodes was ranging from 5 to 15, the number of leaf nodes from 4 to 10 (the maximum depth ranges from 3 to 5 by default). Obviously, the number of leaf nodes and the depth of the tree determine the number of the rules that will be generated and their complexity. Our aim was not to investigate the trade-off between accuracy and complexity of the generated trees.

		Without in-control data (A)		With in-control data (B)	
Representation	Algorithm	Mean	St. Deviation	Mean	St. Deviation
RAW	CHAID	78.90	5.09	66.20	1.70
	CRT	95.75	0.49	75.23	2.48
	QUEST	84.62	18.93	62.88	8.60
SAX (α=6)	CHAID	78.00	2.83	69.07	1.30
	CRT	81.18	2.17	65.72	0.99
	QUEST	54.87	2.31	48.40	0.86
SAX (α=8)	CHAID	80.51	2.32	70.62	1.60
	CRT	84.81	1.60	73.22	0.82
	QUEST	62.94	9.93	68.90	6.31
SAX (α=10)	CHAID	86.26	2.22	76.24	1.34
	CRT	88.06	0.77	75.63	1.76
	QUEST	52.67	3.57	68.53	7.78
FB	CHAID	80.64	2.34	64.59	0.93
	CRT	99.90	0.08	88.54	0.74
	QUEST	77.85	13.63	69.42	10.87

Table 1. Accuracy rates

By far, the most accurate results were obtained with the FB representation and the CRT algorithm on both datasets A (99.9%) and B (88.54%). As it was expected, accuracy rates for datasets with in-control data not included were higher than those for datasets with in-control data included except from the case where SAX (α=8) and SAX (α=10) with QUEST algorithm were applied. The QUEST algorithm shows a noticeable instability in results within each experiment having a standard deviation that reaches 18.93 units, when in all other cases the corresponding standard deviations are smaller than 3 units. Moreover, it gives the “worst” results in most of the cases. One exception is when feature-based representation is applied on dataset B (69.42%), but still it is far behind the best outcome (88.54%). A second exception is when QUEST is applied on raw data of A (84.62%). In both datasets, CRT performs considerably better than CHAID with either raw data or feature-based representation, but the results are comparable when SAX representation is used. Both of these two algorithms show an improving performance with SAX representations as the alphabet size increases from 6 to 10 by approximately 10% in both datasets. The decision trees generated from raw data provide comparable accuracy rates with the trees generated by transformed data. The role representation schemes play in classification within SPC context will be further investigated in future work.

Focusing on SAX (α=10) and FB representations with CHAID and CRT algorithms, some interesting observations were made based on the analytical tables of classification rates from each experiment (these tables are not presented here). On the one hand, FB representation is almost perfect when applied with CRT, but problems appear when applied with CHAID. First, shift patterns are misclassified mainly as trends and secondary as natural or cyclic patterns. The classification rate for shift patterns (upward and downward) ranges from 44% to 98%. Similar problems appear in distinguishing natural from cyclic patterns. When in-control data is present, the major problem is in identifying cyclic patterns, since the accuracy rate ranges from 0% to 17%. CRT shows a little difficulty with distinguishing shifts from trends. On the other hand, the SAX (α=10) representation seems to have a major problem in distinguishing natural from cyclic patterns. In particular, the classification accuracy for natural patterns ranges from 19% to 70%

when CHAID algorithm applied, whereas, the corresponding range for CRT algorithm is from 36% to 47%. Even worse results are obtained in the presence of in-control data for both algorithms: 0%-67% for CHAID and 0%-30% for CRT.

As mentioned above, the accuracy rates reported are affected by the desired complexity of the generated decision trees and, thus, they cannot be compared with other experiments in the literature. In addition to that, these results have not been statistically validated. Nevertheless, they provide some indications of the possible strengths and weaknesses of the various representation schemes and algorithms. Finally, these results indicate that Time Series Data Mining approaches may be comparable with other approaches such Neural Networks.

5. CONCLUSION

In this paper, the possible contribution of Time Series Data Mining in control chart pattern recognition has been investigated. We approached this problem by selecting fairly simple options, among a wealth of TSDM representations, similarity measures and data mining methods. Feature-based and Symbolic Aggregate approXimation representations, although different to each other, are both easy to be implemented. Moreover, in conjunction with Decision Trees for the classification task, they can produce understandable rules that may facilitate the procedure of interpreting control charts and taking a corrective action as fast as possible. The experiments were performed on synthetically generated datasets, taking into account many different parameters such as noise, magnitude of deviations and presence of in-control data. The evaluation was based on the classification accuracy rates and comparisons were made among three different algorithms (CHAID, CRT and QUEST) and between two different datasets (with and without in-control data present in training and testing examples). Results showed that although the generated models were simple the accuracy rates were adequately high, especially when feature-based representation was implemented and CRT algorithm was utilized for classification. SAX representation performed significantly better when the alphabet size was ten. Future work will focus on two areas. First, we will conduct more experiments with respect to different representations, data mining methods and evaluation criteria (e.g., ARL type of indices). Second, we will take into account more real-life parameters relevant to Statistical Process Control in order to adjust and consolidate an appropriate TSDM approach.

References

[1] Cheng C.S. (1995), "A Multi-layer Neural Network Model for Detecting Changes in the Process Mean", *Computers & Industrial Engineering*, 28(1), pp. 51-61.

[2] Cheng C.S. (1997), "A Neural Network Approach for the Analysis of Control Chart Patterns", *International Journal of Production Research*, 35(3), pp. 667-697.

[3] Daw, C. S., Finney, C. E. A. & Tracy, E. R. (2003), "A review of symbolic analysis of experimental data", *Review of Scientific Instruments*, 74(2), pp.915-930.

[4] Deming, W. E. (1986), "Out of the crisis", Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

[5] Deming, W. E. (1982), "Quality Productivity and Competitive Position", Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

[6] Evans J. R. and Lindsay W. M. (1988), "A Framework for Expert System Development in Statistical Quality Control", *Computers & Industrial Engineering*, 14(3), pp. 335-343.

[7] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From Data Mining to Knowledge Discovery: An Overview", In Advances in knowledge Discovery and Data Mining, AAAI / MIT Press, pp. 1-30.

[8] Guh R.S. (2005), "A Hybrid Learning Based Model for On-line Detection and Analysis of Control Chart Patterns", *Computers & Industrial Engineering*, 49(1), pp. 35-62.

[9] Gunopulos D., Das G. (2001), "Time Series Similarity Measures and Time Series Indexing", in Proc. of 2001 ACM SIGMOD international conference on Management of Data, Santa Barbara, California, USA, p.624.

[10] Hosni Y. A. and Elshennawy A. K. (1988), "Quality Control & Inspection: Knowledge-based Quality Control System", *Computers & Industrial Engineering*, 15(1-4), pp. 331-337.

[11] Hwarng H. B. and Hubele N. F. (1993a), "X-bar Control Chart Pattern Identification Through Efficient Off-line Neural Network Training", IIE Transactions, 25, pp. 27-39.

[12] Hwarng H. B. and Hubele N. F. (1993b), "Back-propagation Pattern Recognizers for X-bar Control Charts: Methodology and Performance", *Computers & Industrial Engineering*, 24(2), pp. 219-235.

[13] Keogh, E. & Kasetty, S. (2002), "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration", in Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, pp 102-111.

[14] Lin J., Keogh E., Lonardi S., and Chiu B., (2003), "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms", in Proc. of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA., pp 2-11. 2003

[15] Montgomery, D. C. (1996), "Introduction to statistical quality control", 3rd ed, New York: Wiley.

[16] Murthy, S.K. (1998), "Automatic Construction of decision trees from data: A multidisciplinary survey", *Data Mining and Knowledge Discovery*, 2(4), pp. 345-389.

[17] Nanopoulos A., Alcock R. and Manolopoulos Y. (2001), "Feature-based Classification of Time-series Data", Information Processing and Technology, pp. 49-61.

[18] Nelson, L.S. (1984). "The Shewhart Control Chart – Tests for Special Causes", *Journal of Quality Technology*, 16(4), pp. 237-239.

[19] Nelson, L. S. (1985), "Interpreting Shewhart x-bar control charts", *Journal of Quality Technology*, 17(2), 114–116.

[20] Pugh G. A. (1989), "Synthetic Neural Networks for Process Control", *Computers & Industrial Engineering*, 17(1-4), pp. 24-26.

[21] Shewhart, W.A. (1931), "Economic Control of Quality of Manufactured Product", Van Nostrand: reprinted American Society for Quality Control, 1980, Milwaukee, WI.

[22] Swift J. A. and Mize J. H. (1995), "Out-of-Control Pattern Recognition and Analysis for Quality Control Charts Using Lisp-based Systems", *Computers & Industrial Engineering*, 28(1), pp. 81-91.

[23] Western Electric, (1956), "Statistical Quality Control Handbook", Princeton, NJ: AT&T.

[24] Zorriassatine F. and Tannock J. D. T. (1998), "A Review of Neural Networks for Statistical Process Control", *Journal of Intelligent Manufacturing*, 9(3), pp. 209-224.