Feature selection for SPC chart pattern recognition using fractional factorial experimental design

Adnan Hassan, Mohd. Shariff Nabi Baksh,
Awaluddin M. Shaharoun and Hishamuddin Jamaluddin

Faculty of Mechanical Engineering, Universiti Teknologi Malaysia
81310 UTM Skudai, Johor Bahru
Malaysia

Abstract
Feature selection is one of the important steps in designing a pattern recognizer. This paper presents a study to select a minimal set of statistical features for SPC chart pattern recognition using the fractional factorial experimental design. A resolution IV design was used to identify the significant features from a list of ten possible candidates to represent the input data streams. Further judgment was adopted to arrive at the final selection in the light of some ambiguities among confounded two-factor interactions. The final six selected features set comprising, autocorrelation, cusum, mean, standard deviation, mean-square value, and skewness as the input vector resulted in an average correct classification rate of 97.1% and standard deviation of 0.878. The methodology adopted in this study could be applied to other feature selection problems beside for SPC chart pattern recognition.

Keywords: Feature selection, Pattern recognition, Statistical process control

1. Introduction

Statistical process control (SPC) charts are widely used for 'listening to the voice of the process' [1]. They were introduced by Shewhart in 1924 and remain among the most important SPC tools. The key feature of control charts is the provision of the method to differentiate whether a particular processes is operating within a statistically stable or an unstable state. Unstable processes may produce distinct time series patterns. Identification of these patterns coupled with engineering knowledge of the process would lead to focused diagnosis and troubleshooting.

Developments in computing technology have motivated researchers to explore the use of artificial intelligence technologies such as expert systems, artificial neural network (ANN) and fuzzy sets to automatically and intelligently recognise control chart patterns (CCP). Input data representation is an important step in the design and application of a pattern recognizer. Input data in their original format are normally large in dimension, and their key properties are often hidden and contaminated with noise. Dimensionality reduction can be achieved through extraction of statistical features from the data streams (feature extraction). However, using all candidate features to represent the input data may not necessarily result in better discrimination capability. There may possibly be interactions between the features. The presence of excessive input features can burden the ANN training process, while using too few features may result in insufficient representation. Thus, there is a need to identify the right mix of the statistical features. This exercise is commonly referred to as feature selection. Theodoridis and Koutroumbas [2] noted that the feature selection step is very crucial in designing a pattern recogniser. An appropriate mix of relevant features is expected to have crucial impact on the learning and generalization performance [3]. A smaller ANN size can lead to faster training and generally more effective and efficient recognition.

Most of the existing SPC pattern recognition models in the literature use normalized raw data as the input vector to the recognizer. These data representations normally produce large ANN structures and are not very effective and efficient for complicated recognition problems [4]. These drawbacks can be overcome with the use of features for representing data as demonstrated in pattern recognition applications for hand written [5], characters [6], and grain grading [7] among others.

Very limited works have been reported using features extracted from SPC chart signals for input data representation. Pham and Wani [8] represented input vectors using nine geometric features, namely, slope, number of mean crossings, number of least-square line crossings, cyclic membership, average slope of the line segments, slope difference, and three different measures for area. The aspect of feature selection was not included in their studies. The objective of this paper is to propose a minimal set of statistical features for SPC chart pattern recognition selected using the fractional factorial experimental design. The rest of this paper is organized as follows: Section 2 discusses the sample SPC chart patterns and statistical features, Section 3 describes the ANN pattern recognizer, Section 4 discusses the feature selection and Section 5 concludes the paper.

2. Sample SPC chart patterns and statistical features

This study focuses on recognition of six control chart patterns plotted on the Shewhart *X-bar* chart, namely, random, shift-up, shift down, trend-up, trend-down, and cyclic. Since a large amount of samples were required for the recognizers' training and they were not economically available, simulated data was used. The parameters used for simulating the data streams are given in Table 1. Each data stream consisted of 20 subgroups averages of time sequence data sampled with a rational subgroup size of 5.

Table 1. Parameters for simulating SPC chart patterns

Pattern Type	Parameters (in terms of σ)
Linear trend-up	gradient: 0.015 to 0.025
Linear trend-down	gradient: -0.025 to -0.015
Sudden shift-up	shift magnitude: 0.7 to 2.5
Sudden shift-down	shift magnitude: -2.5 to -0.7
Cyclic	amplitude: 0.5 to 2.5
Random (stable process)	mean = 0; standard deviation =1

The values of the parameters varied randomly in a uniform manner between the limits shown. Random noise of 1/3 σ was added to all unstable patterns. These parameters were chosen to keep the patterns within the control limits. To achieve dimensionality reduction, the following statistical features were extracted from the sample patterns: mean (MEAN), mean-square value (MSV), median (MEDIAN), standard deviation (STD), range (RANGE), maximum cusum (CUS), skewness (SKEW), kurtosis (KUR), slope (SLOPE), and average autocorrelations of lag 1 and 2 (ACOR).

3. The ANN pattern recognizer

Various researchers have shown the feasibility of ANN for SPC chart pattern recognition. The present study adopted the ANN as a tool in developing the recognition system, based on Multilayer Perceptions (MLP) architecture. The number of input nodes was set according to the actual number of statistical features used as the input vector. It falls between 2 to 10 nodes corresponding to the number of features included in the investigation. The number of nodes in the hidden layer was set to 6 which was chosen based on trial and error. The number of output nodes was set to 6 which is the number of pattern classes. Back-propagation with adaptive learning rate and momentum was adopted as the training algorithm. It was assumed that all patterns were fully developed when they appeared within the recognition window. The extracted statistical features were then normalized such that their values would fall within [-1, 1].

4. Feature selection using fractional factorial experimental design

Keeping the number of features as small as possible and with correct combination is in line with the desire to produce recognizers with good generalization capabilities [3]. Fractional factorial experimental designs (FFED) were used to identify features that had substantial effect on the recognizers' performance. FFED are statistical quality engineering techniques which are widely used in screening experiments for product and process design. Detailed discussion on FFED can be found in Montgomery [9].

4.1 Preliminary feature screening

In the preliminary feature screening, a two-level 2_{IV}^{10-5} fractional factorial design was used. This design requires only 32 experimental runs and provides information at resolution IV. Table 2 provides the design matrix for the 2_{IV}^{10-5} fractional factorial design, the feature-column assignment and the experimental results. Since no prior knowledge was available on possible interactions, the features were assigned randomly to the columns. The factor

Table 2. Design matrix for 2_{IV}^{10-5} fractional factorial and experimental results

RUN	ACOR (A)	SLOPE (B)	CUS (C)	MEAN (D)	MSV (E)	STD (F)	SKEW (G)	KUR (H)	MED (J)	RANGE (K)	% Correct Recognition
1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	58.1
2	+1	-1	-1	-1	-1	-1	-1	-1	-1	+1	37.3
3	-1	+1	-1	-1	-1	-1	-1	-1	+1	-1	51.6
4	+1	+1	-1	-1	-1	+1	+1	+1	-1	-1	84.2
5	-1	-1	+1	-1	-1	-1	-1	+1	-1	-1	48.9
6	+1	+1	-1	-1	-1	+1	+1	-1	+1	-1	93.8
7	-1	+1	+1	-1	-1	+1	+1	-1	-1	+1	84.9
8	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1	73.7
9	-1	-1	-1	+1	-1	-1	+1	-1	-1	-1	67.6
10	+1	-1	-1	+1	-1	+1	-1	+1	+1	-1	80.3
11	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	70.6
12	+1	+1	-1	+1	-1	-1	+1	-1	+1	+1	93.6
13	-1	-1	+1	+1	-1	+1	-1	-1	+1	+1	83.9
14	+1	-1	+1	+1	-1	-1	+1	+1	-1	+1	84.6
15	-1	+1	+1	+1	-1	-1	+1	+1	+1	-1	70.9
16	+1	+1	+1	+1	-1	+1	-1	-1	-1	-1	82.5
17	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	76.8
18	+1	-1	-1	-1	+1	-1	+1	+1	+1	-1	88.8
19	-1	+1	-1	-1	+1	-1	+1	+1	-1	+1	75.1
20	+1	+1	-1	-1	+1	+1	-1	-1	+1	+1	87.2
21	-1	-1	+1	-1	+1	-1	+1	-1	+1	+1	71.3
22	+1	-1	+1	-1	+1	+1	-1	+1	-1	+1	76.4
23	-1	+1	+1	-1	+1	+1	-1	+1	+1	-1	69.7
24	+1	+1	+1	-1	+1	-1	+1	-1	-1	-1	92.8
25	-1	-1	-1	+1	+1	-1	-1	+1	+1	+1	63
26	+1	-1	-1	+1	+1	+1	+1	-1	-1	+1	96.4
27	-1	+1	-1	+1	+1	+1	+1	-1	+1	-1	74.3
28	+1	+1	-1	+1	+1	-1	-1	+1	-1	-1	73
29	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1	78.9
30	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1	90.1
31	-1	+1	+1	+1	+1	-1	-1	-1	-1	+1	71.5
32	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	92.3

(feature) in each column is independent and free from being confounded with other factors.

The performance of the recognitions was measured based on percentage of correct recognition. A total of 2160 and 720 sample patterns were used in the training and recall phases of each of the screening run, respectively. Additional 720 sample patterns were used for in-training validation for early stopping to avoid over fitting during training. Each run was replicated using 10 different data sets in order to minimize the random error.

The recognizer responses as given in Table 2 indicate that when all the features were included (run 32), the mean performance was not the best (92.3%). Better or somewhat similar performance results were achievable with a reduced number of features as indicated in trial run 6 (93.8%), run 12 (93.6%), run 24 (92.8%) and run 26 (96.4%). This suggests that there were effects of possible interactions between the features. It was necessary to identify the features to be retained and those to be omitted.

The experimental results (32 runs × 10 replicates) were analyzed using the *MINITAB* statistical package for ANOVA and interaction effects. Table 3 summarizes these tentative significant effects along with the 2 factor interactions and their aliases. Interaction here means the effect of the first feature depends on the level of the second feature (synergistic effect). The symbol '*' is used to indicate the pair of interacting features.

As noted earlier, this study used a resolution IV design for running the experiments. The source of ambiguity in this design was that all the 2-factor interactions were confounded with other 2-factor interactions. One possible approach to resolve the confounding was by using resolution V design that would free the two factor interactions from other two factor interactions. However, the resolution V with 10 features was not attractive from economic considerations since it would require 128 runs compared with only 32 runs in the present study. Jones [10] noted that, often, engineering judgement or some clarification experiments be conducted to clarify such ambiguity economically. This approach was adopted in this study.

The list for tentative significant main effects listed in Table 3 does not include features MEDIAN, RANGE and SLOPE. Possibly, features MEDIAN and RANGE were duplicates to features MEAN and STD respectively. However, the role of feature

SLOPE was still unclear at this stage. By assuming these three features also did not contribute significantly in any of the two factor interactions (or their aliases), led to elimination of related “unimportant” interactions. Thus, the tentative significant two factor interactions listed in Table 3 were reduced to the following: MEAN*MSV, STD*SKEW, CUS*KUR, MEAN*SKEW, MSV*STD, ACOR*SKEW, CUS*MSV, and STD*KUR. It should be reiterated here that there is a risk of reaching false conclusions when making such assumptions. However, one can minimize such risk by conducting clarification and confirmation runs [10].

The interaction subplots in Figure 1 were scrutinized to evaluate the sensitivity of the remaining main effects and their interactions. As indicated by the dotted response line in the ACOR*SKEW subplot, the recognizer performance was significantly improved when both features Autocorrelation and Skewness were included. Excluding any of these features severely reduced the recognizer performance. This subplot also indicates that by including feature autocorrelation, the response became more sensitive to changes in skewness. This may not be good if the feature autocorrelation was contaminated with noise. However, it was decided that both autocorrelation and skewness features should be included since the improvement in the recognition performance was very significant.

The subplots for CUS*KUR and STD*KUR suggest that significantly higher recognizer performance could be achieved when the second feature (kurtosis) was excluded. The same subplots also indicate that the inclusion of both features significantly reduced the recognizer’s performance. Since these interactions (CUS*KUR and STD*KUR) were confounded with other 2-factor interactions, the decision to exclude the feature kurtosis was tentative and subject to clarification runs.

Table 3. Tentative significant main effects and two factor interactions

MAIN EFFECTS	INTERACTIONS	AND	ALIASES
ACOR	ACOR*MED	+	SLOPE*RANGE
SKEW	MEAN*MSV	+	STD*SKEW
STD	SLOPE*MEAN	+	SKEW*MED
MSV	ACOR*SKEW	+	MEAN*RANGE
MEAN	CUS*MSV	+	STD*KUR
CUS	ACOR* RANGE	+	SLOPE*MED + CUS*KUR + MEAN*SKEW + MSV*STD
KUR			

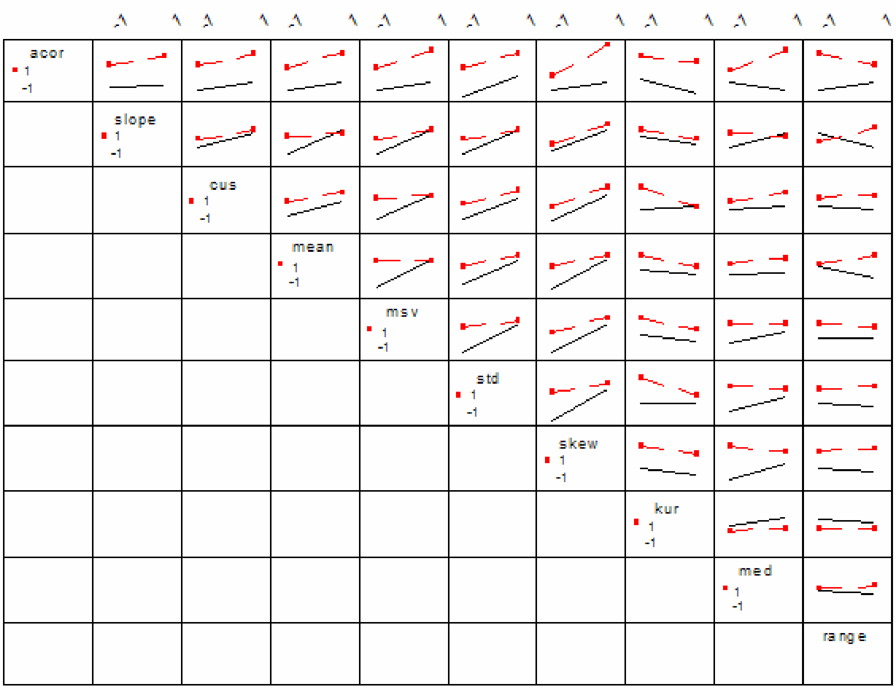

Figure 1: Interaction subplots

The response lines in the subplots for interactions MEAN*MSV, MEAN*SKEW, STD*SKEW, MSV*STD, CUS*MSV indicate that the inclusion of the first features (row features) only marginally improved the recognizers' performance while the second features (column features) were at "+1" settings. The solid response lines for these subplots also indicate that the exclusion of the respective first feature (row) resulted in the recognizer becoming more sensitive to changes in the second (column) features. This may be undesirable if the change in the level of second features was contributed by noise. At this stage, it was more appropriate to maintain features MEAN, MSV, SKEW, STD and CUS since the evidence available did not warrant their exclusion. Thus, clarification and confirmation runs were necessary to clarify the ambiguity and minimize the risk of false conclusions.

4.2 Clarification and Confirmation Experiments

Before proceeding with clarification and confirmation experiments, it would be useful to fit the previous experimental results to a reduced linear regression model. A reduced model was formulated by screening out the "unimportant" features, namely, median, range, slope, kurtosis and their related interactions as identified in the earlier analysis. Table 4 provides the results of this fitting and Table 5 gives its corresponding analysis of variance.

Table 5 shows that the reduced model fitted satisfactorily with a small lack of fit error (F_0 = 1.27 and p = 0.193). Features listed in the reduced model would form a reference setting for the clarification and confirmation experiments. Table 6 summarises the settings and experimental results for these experiments. Setting 1 is the reference setting that includes all the tentatively accepted features. Settings 2 through 6 remove one of these features (as marked '-1' in the respectively column) for each of the subsequent experiments. Finally, setting 7 adds feature slope to the reference setting. This setting attempted to clarify the effect of feature slope that was assumed "insignificant" at the early stage of the analysis. Each of the above setting was evaluated using 720 test patterns (120 patterns for each type). The evaluation for each setting was replicated using 10 different test sets.

The reference setting 1 resulted in 97.10% for average correct recognition with standard deviation of 0.88. Its corresponding predicted value based on the reduced model was 96.54%. Table 6 shows that when omitting any of the tentatively accepted features, the performance of the recogniser was significantly reduced. Similarly, the recogniser's performance deteriorated (92.55%) when feature slope was added to the reference setting. Thus, the minimal feature set should comprise of the following features: autocorrelation, cusum, mean, mean-square-value, standard deviation, and skewness.

Table 4. A reduced model for the recogniser's performance

Term	Estimated Effect	Regression Coefficient	*t*	*p*-value
Constant		76.378	63.01	0.000
ACOR	13.101	6.551	5.40	0.000
CUS	5.483	2.742	2.26	0.024
MEAN	6.435	3.218	2.65	0.008
MSV	6.953	3.476	2.87	0.004
STD	8.546	4.273	3.53	0.000
SKEW	10.702	5.351	4.41	0.000
ACOR*SKEW	5.083	2.541	2.10	0.037
CUS*MSV	-4.467	-2.233	-1.84	0.066
MEAN*MSV	-6.279	-3.140	-2.59	0.010
MEAN*SKEW	-5.244	-2.622	-2.16	0.031

Table 5. ANOVA for the recogniser's performance (reduced model)

Source of Variation	Deg. of Freedom	Seq. Sum of Squares	Adj. Sum of Squares	Adj. Mean Square	F_0	*p-value*
Main Effects	6	38323	38323	6387.2	13.59	0.000
2-Way Interactions	4	9018	9018	2254.4	4.80	0.001
Residual Error	309	145280	145280	470.2		
Lack of Fit	21	12327	12327	587	1.27	0.193
Pure Error	288	132953	132953	461.6		
Total	319	192620				

Table 6. Settings and results of confirmation/clarification experiments

Settings	Features							Experimental Results* (% Correct)		Predicted (% Correct)
	ACOR	CUS	MEAN	MSV	STD	SKEW	SLOPE	Mean	Std. Dev.	Mean
1	+1	+1	+1	+1	+1	+1	Nil	97.10	0.88	96.54
2	+1	+1	+1	+1	-1	+1	Nil	90.35	7.30	
3	+1	+1	-1	+1	+1	+1	Nil	91.04	10.20	
4	+1	+1	+1	-1	+1	+1	Nil	93.38	7.02	
5	+1	-1	+1	+1	+1	+1	Nil	93.44	6.80	
6	+1	+1	+1	+1	+1	-1	Nil	81.26	27.38	
7	+1	+1	+1	+1	+1	+1	+1	92.55	13.39	

* Based on 10 replicates, each comprising a total of 720 test samples (120 for each type)

5. Conclusions

A final feature set comprising autocorrelation, cusum, mean, mean-square-value, standard deviation, and skewness is proposed. Rejection of features median and range is possibly due to their being redundant, given that other features are included. The central tendency and spread characteristics of the patterns may have been captured sufficiently by the statistics mean and standard deviation respectively. The results did not favor the inclusion of the feature kurtosis, possibly due to its role in providing information on the fourth moment. As such, kurtosis was relatively less robust to noise. The results also did not provide sufficient evidence to support the inclusion of the feature slope. This is possibly due to the limitation of resolution IV fractional factorial design, which had some ambiguities among 2-way interaction. However, the approximate method adopted in this study, that is the fractional factorial design coupled with judgment seems to be sufficient for this feature screening.

This study has shown that fractional factorial experimental design is a viable approach for feature selection. Its comparison with other feature selection techniques can be an opportunity for further study. Although an ANN-based pattern recognizer for SPC chart pattern is the particular application presented here, the methodology adopted in this study can be applied to other feature selection problems. Currently, we are extending this work to recognize control chart patterns as they are developing [11].

Acknowledgments

The authors thank Research Management Center, Universiti Teknologi Malaysia (UTM) and the Ministry of Science, Technology and Innovation (MOSTI), Malaysia for IRPA research grant Vot. No. 74227.

References

[1] Oakland JS. Statistical Process Control. Butterworth-Heinemann. Oxford (1996).

[2] Theodoridis S and Koutroumbas K. Pattern Recognition Systems. Academic Press, San Diego. (1998).

[3] Battiti R. Using Mutual Information for Selecting Features in Supervised Neural Net Learning. IEEE Transactions on Neural Networks. Vol. 5, No. 4 (1994) pp. 537-550.

[4] Hassan A, Baksh MSN, Shaharoum AM and Jamaluddin H. Improved SPC Chart Pattern Recognition Using Statistical Features. International Journal of Production Research. Vol. 41, No. 7. (2003) pp. 1587 – 1603.

[5] Zeki AA and Zakaria MS. New Primitive to Reduce the Effect of Noise for Handwritten Features Extraction. IEEE 2000 Tencon Proceedings: Intelligent Systems and Technologies for the New Millennium. (2000) pp. 24-27.

[6] Amin A. Recognition of Printed Arabic Text Based on Global Features and Decision Tree Learning Techniques. Pattern Recognition. 33, (2000) pp. 1309-1323.

[7] Utku H. Application of the Feature Selection Method to Discriminate Digitized Wheat Varieties. Journal of Food Engineering, 46, (2000) pp. 211-216.

[8] Pham DT and Wani MA. Feature-based Control Chart Pattern Recognition. Int. J. Prod. Res. Vol. 35, No.7 (1997) pp. 1875-1890.

[9] Montgomery DC. Design and Analysis of Experiments. 5^{th} edn., (2001) John Wiley & Sons, New York.

[10] Jones B. Design of Experiments. In Quality Engineering Handbook (Pyzdek, T. and Berger, R.W., eds.) (1991). Marcel Dekker, New York, pp. 329-387.

[11] Hassan A and Baksh MSN. Recognition of Developing SPC Chart Patterns Using ANN and Stability Tests. International Symposium on Bio-inspired Computing. 5-7 September 2005, Johor Bahru, Malaysia.