An Effective and Novel Neural Network Ensemble for Shift Pattern Detection in Control Charts

Mahmoud Barghash

IE Department, The University of Jordan, Amman 11942, Jordan

Correspondence should be addressed to Mahmoud Barghash; mabargha@ju.edu.jo

Received 7 May 2015; Accepted 6 July 2015

Academic Editor: José David Martín-Guerrero

Pattern recognition in control charts is critical to make a balance between discovering faults as early as possible and reducing the number of false alarms. This work is devoted to designing a multistage neural network ensemble that achieves this balance which reduces rework and scrape without reducing productivity. The ensemble under focus is composed of a series of neural network stages and a series of decision points. Initially, this work compared using multidecision points and single-decision point on the performance of the ANN which showed that multidecision points are highly preferable to single-decision points. This work also tested the effect of population percentages on the ANN and used this to optimize the ANN’s performance. Also this work used optimized and nonoptimized ANNs in an ensemble and proved that using nonoptimized ANN may reduce the performance of the ensemble. The ensemble that used only optimized ANNs has improved performance over individual ANNs and three-sigma level rule. In that respect using the designed ensemble can help in reducing the number of false stops and increasing productivity. It also can be used to discover even small shifts in the mean as early as possible.

1. Introduction

Pattern recognition in control charts is extremely important for increasing productivity and for better customer satisfaction. Manufacturing process is of random nature; however certain aspects are not random such as broken heater, worn bearings, unacceptable raw materials, worn coils, higher than needed temperature, or other process parameters. These are called assignable causes. Assignable causes (real process uncontrollable changes) can affect the process quality and must be discovered as early as possible while unassignable random causes (normal process variations) must be ignored. In case the process is allowed to continue working under assignable causes effect, quality deteriorates and more defects or defectives are produced. This case is called type II error. If the process is stopped when the process is still under normal variation (this is a false alarm) then the productivity is reduced. This is called type I error. Both errors are current in practice. Successful process control is a technique that minimizes both errors. There is usually a compromise between these two errors; that is, if we try to reduce type II error, that is, to make sure that more nonnormal causes are discovered, then more normal variations are mistaken as nonnormal and type I error is increased. If we try to reduce type I error, that is, to make sure that less normal variations are mistakenly classified as assignable, then more assignable causes are undiscovered and type II error is increased. Early discovery of faults is a subject of a large sum of current research. This is usually combined with online monitoring when this is related to manufacturing processes. Some examples of this failure monitoring are induction motor failure discovery using current signal [1], bearing fault detection [2], metal processing sensor fault diagnostics [3], and so forth. In all, the decision should be made of whether the process is subject to real change (assignable) or whether the process is still under normal causes. Figure 1 maps the generated patterns from normal and faulty processes against the decision criterion for failure. Area B shows patterns that can be generated by both faulty and normal processes and represents a challenge for the researchers. Areas A and E are successfully classified as faulty (A) when the process is faulty and not faulty (E) when it is not faulty. Areas D and C are the main decision mistakes leading to type II error (D) whereby the process is classified as normal...
while the process is actually faulty and type I error whereby the process is classified as faulty while it is actually normal.

The main objective of statistical process control is to develop techniques that can be used to reduce type I and type II errors. The early attempts were based on heuristic rules of thumb. Control charts were proposed by Shewhart in the early 1920s which was improved through heuristics and zone testing by Western Electric rules in 1958 [4]. Page in 1958 had developed the Cumulative SUM (CUSUM) control chart and Roberts in 1959 had proposed the Exponentially Weighted Moving Average (EWMA) control charts to increase the sensitivity to discover even small changes in the processes. Also Lucas in 1973 developed a V-mask for the CUSUM control chart to improve its abilities. The sole purpose of these traditional tools is to improve the ability to discover changes in the process. However, while using these tools, the compromise between type I and type II errors still holds. This is the reason why many researchers turned to Artificial Intelligence [5]. Several research works have been conducted in pattern recognition in control charts using Artificial Intelligence including Artificial Neural Networks (ANN) [6–12], support vector machine (SVM) [13], rule-based [14], expert system [15, 16], neurofuzzy [17], multistage neural networks [18], feature selection [19, 20], fuzzy [21], ART [22], learning vector quantization [23], neural expert [24], ensemble based techniques [5], and efficient features combined with optimized ANN [25].

Recently ensemble based techniques have seized more attention by researchers. Several research works have shown improved results when using ensemble based techniques. Ensemble based techniques identified both mean and variance shifts [26], autoregressive coefficient-invariant control chart pattern recognition model for online recognition of seven typical types of unnatural control chart patterns [27], both raw data and statistical features extracted combined with an ANN ensemble, which are used for the recognition which provides for better performance [28], a selective NN ensemble of selected NNs to classify source(s) of out-of-control signals in multivariate processes [29], identifying the source(s) of process mean shifts in multivariate control charts using an improved particle swarm optimisation with simulated annealing-based selective multiclass support vector machines ensemble [30–32] and diagnosing and classifying out-of-control signals in a bivariate process using self-organization map neural network [33], and percent correct classification for a committee of specialized recognizers [34]. While previous literature work concentrated on developing new Artificial Intelligence techniques or training algorithms, this research work focused on optimizing the performance and stability of the ensemble construction to achieve low false alarms and quick discovery of even the smallest shift in the process mean. This research work shows that different ensemble construction can give highly different results and thus emphasized the need for ensemble tuning or optimization.

Shift patterns are some of the most common assignable cause patterns. These occur when the process mean changes due to assignable causes. This is usually reflected in product characteristic changes such as size, shape, or other quality parameters. A usual measure for the success of SPC system is the average run length where the number of signal samples before an alarm is sounded and the process is out of control. Figure 2 shows a sample ARL chart which is generated based on the basic benchmark traditional X-bar. The x-axis is the shift size. There is no shift at the origin or point zero. When the shift size is small, close to the origin, there will be some mix-up between shift and no shift case, while at large shift case, far from the origin, there is usually no mix-up with no shift case. The ARL at zero shift size is close to 373; that is, on average, a false alarm is issued every 373 data points. For the case of small shift 0.5 standardized shift (the system mistakenly classifies this as no shift for a large number of samples), nearly 150 data points are passed before the shift is discovered. A better performing system may reduce the number of false alarms while keeping the 0.5 standard shift at 150, or keeping the ARL at zero shift to 373 while reducing the ARL for 0.5 sigma to less than 150, or it can increase the ARL at zero shift and reduces the ARL at 0.5 sigma to less than 150.

This work is a based on previous work (Barghash and Santarisi [35] and Barghash [5]) which showed that extraordinary balance between type I and type II error can be achieved through the use of optimized ANN ensembles. This work shows that compromise between type I and type II error can be greatly improved through proper design of the population percentages and using ensembles. This work also tests the effect of the population of percentages on type I and type II
errors in an attempt to optimize the ANN. This work also tests the use of optimized and nonoptimized ANN for the purpose of detecting the shift pattern which is one of the most common assignable causes in the industry.

2. Artificial Neural Network Ensemble

2.1. Artificial Neural Network Training. A neural network is composed of several interconnected neurons with the following activation function for each neuron [36]:

\[
\text{output} = \varphi \left( \sum_{i=1}^{p} w_{ij} x_i - \theta \right),
\]

where \( \varphi \) is the neuron function, \( w_{ij} \) is the neuron \( i \)th weight, \( x_i \) is the \( i \)th input to the neuron, \( \theta \) is the constant threshold value, and \( p \) is the number of inputs to the neuron.

Neural networks have high interconnectivity, where all the neurons are arranged in layers. The outputs of the first layer are the inputs for the second layer as shown in Figure 3. The first layer is the input layer and the final layer is output layer. The layers in between are hidden layers [6].

Training of the neural network is done using backpropagation where the error value and the local gradient of the error are used to estimate the new value of the weights and biases:

\[
\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n),
\]

where \( \delta_j(n) \) is the local gradient for the error in output neuron \( j \) at iteration \( n \). \( y_i(n) \) is the output of the network at neuron \( i \) at iteration \( n \). \( n \) is the iteration number [15].

Several types of neuron activation functions from MathWorks Matlab 7.0 were used including the Logsig sigmoid function, Tansig, and Purelin:

\[
\text{Logs}(x) = \frac{1}{1 + e^{-x}}, \\
\text{Purelin}(x) = b + x, \\
\text{Tansig}(x) = \text{sign} \left( \frac{2}{1 + e^{-2x}} - 1 \right).
\]

2.2. Artificial Neural Network Ensemble. Figure 4 shows an ANN ensemble which is a series of ANNs. These ANNs are trained sequentially where the second layer ANN is trained using the output of the first ANN layers and so on.

3. Pattern Generation

Figure 5 shows some sample patterns that are usual for the training of ANNs in pattern recognition in control charts, namely; normal (no assignable cause), shifts of both types upward and downward, cyclic, and trend of both types upward and downward.

Equation (4) is used to generate the no shift and the shift patterns:

\[
\text{Generated pattern } y(i) = \mu + n(t) + d(t),
\]

where \( \mu \) is the process mean. \( n(t) \) is the normal cause of variations which is \( N(0, \sigma) \). \( d(t) \) is the special cause of variations.

For the case of a shift \( d(t) = u \ast s \), where \( u \) is the position of the shift (it can be either 1 or 0) and \( s \) is the value of the shift in the mean.

The following standardization was chosen for the patterns:

\[
Y(i) = \frac{y(i) - \mu}{\sigma},
\]

The final variable \( Y(i) \) is a standardized variable with mean zero and a standard deviation of one for the no shift case. The mean shift values \( s \) can be expressed in a standardized form: as \( -4, -3, -2, -1, 0, 0.5, 1, 2, 3, \) and \( 4 \) or they can be expressed as a multiple of the standard deviation. The \( s \) variable is set to zero for the no shift case.

4. Methodology

Figure 6 shows the proposed methodology. In this work we initially tested the effect of population percentage on the ANN performance and then tested the effect of the ensemble construction on performance.

The basic objective is to reduce both type I error and type II error. For that purpose, the average run length (ARL) is selected as it can reflect type I and type II errors in one chart at various shift levels. ARL also can be easily interpreted economically. The ARL represents the average number of points before a declaration of a nonnormal pattern. If the process is
5. Results and Discussion

The results section considers three main phases or objectives: firstly to test the effect of multidecision points on

normal without a shift ideally it should go to infinity before being stopped. If the pattern is coming from a shifted process then the ARL should ideally be very close to zero.

Preparation of an individual neural network based on the work of Barghash and Santarisi [6] is done by generating shift patterns and normal patterns as shown in Figure 7. A randomized selector is used to perform the pattern selection to construct a population with preset pattern type percentages.
5.1 Phase 1: Stabilization of the ANN Behavior. Neural networks suffer from instabilities when using individual decisions. This instability can be defined as giving an erroneous decision temporarily and then giving the right decision back. This temporary misjudgment by the neural network reduces the ARL appreciably and deteriorates the performance of the neural network. This is solved by adding one or more decision points. If the ANN still insists on its decision in two or more following decisions points, this can be a sign on the existence of a nonnormal behavior. Multidecision point stabilization is tested in accordance with Figure 8, where the case of a single decision and 2 consecutive decision points are tested and the results for these two cases are shown in Figures 9 and 10, respectively. The performance of both ANNs is tested at triggering threshold levels 0.1–0.9. Two clear conclusions can be drawn; firstly, the threshold is a tuning parameter for the performance of the ANN.

And secondly, the unstabilised ANN is far worse than stabilized ANN in terms of the number of false alarms. Thus for the rest of this work, only stabilized ANNs are considered.

5.2 The Effect of Population Percentages. The effect of population percentages is analyzed according to Figure 11 where the ANNs are trained in accordance with different population percentages tested for the ARL. Figures 12 and 13 show the results of a nonoptimized 30% normal and 70% shift pattern and the optimized ANN trained with 75% normal and 25% shift pattern. The nonoptimized ANN ARL is worse than that of the benchmark X-bar chart with three-sigma level rule. For the case of optimised ANN, the ARL for the case zero shift is extremely high while reducing the ARL for the case of standard 0.5 shift level to less than 150.

5.3 Ensemble Results. The basic shape of the generalized suggested ensemble is shown in Figure 14. The first stage is a selected neural network from four main categories: normal pattern detectors, shift pattern detectors, trend pattern
detectors, and cyclic pattern detectors. Then the last two decisions of each neural network are passed to the leader net which is trained to give the decision of whether the pattern is normal or not. A question may arise on the reason for considering patterns other than normal and shift patterns in this work, although only shifts are intended to be discovered. The reason has two sides; the first relates to the fact that these patterns are present in real life and in other research works. Thus, it is imperative that the suggested ensembles be designed to discover and classify all patterns, although this work reports only on its ability to discover shift patterns. The second side relates to the problem that the ability to discover shift patterns might be affected if more patterns are included; thus it seems more logical to include these in any realistic attempt to recognize shift patterns in control charts.

5.3.1. Ensemble 1 Results. Ensemble 1 is formed of 3 ANNs devoted to discovering normal patterns, 2 ANNs for shift patterns, one ANN for trend patterns, and one ANN for cyclic patterns. The last decision point is added as an input as well as the output of the selected ANNs to the leader net. If the output of the leader net is greater than an assigned threshold, the pattern is classified as a shift pattern and the process is stopped. The ARL is then the number of prior decision points. The results obtained from the ensemble are shown in Figure 15. The ARL increases as you select a higher threshold. In general, the performance of this suggested ensemble is not as good as the X-bar chart. For example, the length of the ARL at zero shift size (normal) is shorter than that of the benchmark X-bar 3 sigma which is 373.
5.3.2. Ensemble 2 Results. Figure 16 shows the average run length versus standard shift size for an ensemble with 3 ANN normal detecting patterns, 3 ANN shift detecting patterns, 1 ANN detecting cyclic pattern, and 1 ANN detecting trend pattern. The results of this ensemble are slightly better than those for the X-bar chart 3-sigma level for the case of threshold = 0.9. The ARL for the case of normal pattern (zero shift size) is nearly the same as that of the X-bar chart but the ARL at a shift size of 0.5 is lower than that of the X-bar chart. Thus this ensemble has lower type I and II error.

5.3.3. Ensemble III Results. Figure 17 shows the results for the ensemble with 3 shift pattern detection ANNs and 3 normal detecting ANNs. The average run length of this ensemble is lower in ARL than the X-bar chart for both normal and shift behavior.

5.3.4. Ensemble IV. Figure 18 shows the ensemble results for two normal detecting and two shift detecting neural networks. This ensemble has a behavior at threshold of 0.9 that is comparable to that of the X-bar chart.

5.3.5. Ensemble V. Figure 19 shows the ARL for 1 ANN for shift patterns, 1 ANN for normal patterns, 1 ANN for trend patterns, and 1 ANN for cyclic patterns. This ensemble includes only the best performing ANNs of each category which includes 75% normal patterns and 25% other patterns. The behavior of the ensemble at Th = 0.9 is improved for the case of normal pattern but is worse for other shift sizes. But if we closely observe the case of Th = 0.7, then we can notice...
5.3.6. Ensemble VI. Figure 20 shows the ARL results for the case of 1 normal pattern and one shift pattern which are of optimal performance. It can be noticed that, for Th = 0.9, the ARL at zero shift size is extremely higher than that of the X-bar chart while keeping the ARL for the case of shift patterns lower than those for X-bar chart and the individual chart.

6. Conclusions

This work represents a novel new pioneer work with high implications on fault detection. It showed that population optimised ANN ensemble can have superior performance when compared to the traditional benchmark X-bar chart.

This work tested the effect of using multidecision point on the performance of the ANN, which enhanced appreciably the behavior of the ANN. This work also tested the effect of the population construction and reported that the best performance is noticed when the population construction is 75% for normal and 25% for other patterns. This work also tested several ensembles and concluded that including nonoptimized ANNs may reduce the performance of the ensemble, while including only the best performing patterns gives the best results which increases the ARL for the case of zero shift and reduces the ARL for the case of shift patterns.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This was work was conducted during sabbatical leave funded by The University of Jordan.

References

[1] A. H. Bonnett and G. C. Soukup, “Cause and analysis of stator and rotor failures in three-phase squirrel-cage induction motors,” IEEE Transactions on Industry Applications, vol. 28, no. 4, pp. 921–937, 1992.
[2] H. P. Rodriguez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso, “Application of the Teager-Kaiser energy operator in bearing fault diagnosis,” ISA Transactions, vol. 52, no. 2, pp. 278–284, 2013.
[3] D. Theilliol, H. Noura, D. Sauter, and F. Hamelin, “Sensor fault diagnosis based on energy balance evaluation: application to a metal processing,” ISA Transactions, vol. 45, no. 4, pp. 603–610, 2006.
[4] L. S. Nelson, “The Shewart control chart—tests for special causes,” Journal of Quality Technology, vol. 16, pp. 237–239, 1984.
[5] M. A. Barghash, “A diverse neural network ensemble team for mean shift detection in X-Bar and CUSUM control charts,” Jordan Journal of Mechanical and Industrial Engineering, vol. 5, no. 4, pp. 291–300, 2011.
[6] M. A. Barghash and N. S. Santarisi, “Pattern recognition of control charts using artificial neural networks—analyzing the effect of the training parameters,” Journal of Intelligent Manufacturing, vol. 15, no. 5, pp. 635–644, 2004.
[7] C.-S. Cheng, “A multi-layer neural network model for detecting changes in the process mean,” Computers & Industrial Engineering, vol. 28, no. 1, pp. 51–61, 1995.
[8] C.-S. Cheng, “A neural network approach for the analysis of control chart patterns,” International Journal of Production Research, vol. 35, no. 3, pp. 667–697, 1997.
[9] R.-S. Guh and J. D. T. Tannock, “Recognition of control chart concurrent patterns using a neural network approach,” International Journal of Production Research, vol. 37, no. 8, pp. 1743–1765, 1999.
[10] D. T. Pham and E. Oztemel, “Control chart pattern recognition using neural networks,” Journal of Systems Engineering, vol. 2, no. 4, pp. 256–262, 1992.
[11] G. A. Pugh, “A comparison of neural networks to SPC charts,” Computers & Industrial Engineering, vol. 21, no. 1–4, pp. 253–255, 1991.

[12] N. Santarisi and M. A. Barghash, “Neuro-fuzzy based model for pattern recognition in control charts,” in Proceedings of the 37th International Conference on Computers and Industrial Engineering, pp. 1881–1897, October 2007.

[13] X. Shao, “Recognition of control chart patterns using decision tree of multi-class SVM,” in Advances in Intelligent Systems, vol. 138 of Advances in Intelligent and Soft Computing, pp. 33–41, Springer, Berlin, Germany, 2012.

[14] D. T. Pham and M. A. Wani, “Feature-based control chart pattern recognition,” International Journal of Production Research, vol. 35, no. 7, pp. 1875–1890, 1997.

[15] D. T. Pham and E. Oztemel, “XPC: an on-line expert system for statistical process control,” International Journal of Production Research, vol. 30, no. 12, pp. 2857–2872, 1992.

[16] M. Bag, S. K. Gauri, and S. Chakraborty, “An expert system for control chart pattern recognition,” The International Journal of Advanced Manufacturing Technology, vol. 62, no. 1–4, pp. 291–301, 2012.

[17] S. I. Chang and C. A. Aw, “A neural fuzzy control chart for detecting and classifying process mean shifts,” International Journal of Production Research, vol. 34, no. 8, pp. 2265–2278, 1996.

[18] S. I. Chang and E. S. Ho, “Two-stage neural network approach for process variance change detection and classification,” International Journal of Production Research, vol. 37, no. 7, pp. 1581–1599, 1999.

[19] S. K. Gauri and S. Chakraborty, “Recognition of control chart patterns using improved selection of features,” Computers & Industrial Engineering, vol. 56, no. 4, pp. 1577–1588, 2009.

[20] S. K. Gauri and S. Chakraborty, “Feature-based recognition of control chart patterns,” Computers and Industrial Engineering, vol. 51, no. 4, pp. 726–742, 2006.

[21] M. Gülbay and C. Kahraman, “Development of fuzzy process control charts and fuzzy unnatural pattern analyses,” Computational Statistics & Data Analysis, vol. 51, no. 1, pp. 434–451, 2006.

[22] H. B. Hwarng and C. W. Chong, “Detecting process non-randomness through a fast and cumulative learning ART-based pattern recognizer,” International Journal of Production Research, vol. 33, no. 7, pp. 1817–1833, 1995.

[23] D. T. Pham and E. Oztemel, “Control chart pattern recognition using learning vector quantization networks,” International Journal of Production Research, vol. 32, no. 3, pp. 721–729, 1994.

[24] J. Plummer, “Tighter process control with neural networks,” AI Expert, vol. 8, pp. 49–55, 1993.

[25] V. Ranaee and A. Ebrahimzadeh, “Control chart pattern recognition using an optimized neural network and efficient features,” ISA Transactions, vol. 49, no. 3, pp. 387–393, 2010.

[26] B. Wu and J.-B. Yu, “A neural network ensemble model for on-line monitoring of process mean and variance shifts in correlated processes,” Expert Systems with Applications, vol. 37, no. 6, pp. 4058–4065, 2010.

[27] W.-A. Yang and W. Zhou, “Autoregressive coefficient-invariant control chart pattern recognition in autocorrelated manufacturing processes using neural network ensemble,” Journal of Intelligent Manufacturing, 2013.

[28] J. Yu, L. F. Xi, and B. Wu, “A neural network ensemble approach for the recognition of SPC chart patterns,” in Proceedings of the 3rd International Conference on Natural Computation, vol. 2, pp. 575–579, 2007.

[29] J. B. Yu, L. F. Xi, and X. Zhou, “Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble,” Engineering Applications of Artificial Intelligence, vol. 22, no. 1, pp. 141–152, 2009.

[30] S. Du, J. Lv, and L. Xi, “On-line classifying process mean shifts in multivariate control charts based on multiclass support vector machines,” International Journal of Production Research, vol. 50, no. 22, pp. 6288–6310, 2012.

[31] T.-F. Li, S. Hu, Z.-Y. Wei, and Z.-Q. Liao, “A framework for diagnosing the out-of-control signals in multivariate process using optimized support vector machines,” Mathematical Problems in Engineering, vol. 2013, Article ID 494626, 9 pages, 2013.

[32] W.-A. Yang, “Monitoring and diagnosing of mean shifts in multivariate manufacturing processes using two-level selective ensemble of learning vector quantization neural networks,” Journal of Intelligent Manufacturing, 2013.

[33] J. Yu and L. Xi, “A hybrid learning-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes,” International Journal of Production Research, vol. 47, no. 15, pp. 4077–4108, 2009.

[34] A. Hassan, “Ensemble ANN-based recognizers to improve classification of X-bar control chart patterns,” in Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM ’08), pp. 1996–2000, December 2008.

[35] M. Barghash and N. Santarisi, “A study on the response and stability of artificial neural networks for pattern recognition in control charts,” in Proceedings of the 12th International Conference on Machine Design and Production, Kuşadası, Turkey, September 2006.

[36] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, Englewood Cliffs, NJ, USA, 1994.