LUND UNIVERSITY

Department of Statistics

An Artificial neural network approach for process variance change detection for both small and large shifts

Master Thesis

(15 ECTS)

Spring 2010

Author: Babar Zaman

Supervisor: Prof. Björn Holmquist

Contents

Pages

An Artificial neural network approach for process variance change detection for both small and large shifts

Abstract

Nowadays Statistical process control charts (SPCC) has become powerful tools to monitor process variability of products in manufacturing environments. The artificial neural network (ANN) technique is also nowadays popular to use for monitoring process variability of products as an alternative to SPCC due to its superior performance. This research focus on the ANN technique for monitoring the process variance. We describe the ANN approach to monitor process variance change and shows that it is more efficient than traditional statistical control charts ($R\ charts\ and\ EWMA\ charts$) for both small and large shifts. The performance of the proposed scheme (ANN) is compared to SPCC for samples of size five in terms of average run length to detection of change. The robustness of the proposed scheme (ANN) is also examined.

1. Introduction

These days, to monitor process variability of products in manufacturing environments is a big issue. Statistical process control charts (SPCC) are useful to detect process variability due to specific assignable causes. But artificial neural networks (ANN) are also strongly recommended as alternatives to SPCC to detect process variability due to their superior performance.

1.1. Quality Control charts

The use of Statistical process control charts (SPCC) has become a most powerful technique for detecting the variability and improving quality of the products through the reduction of undesired variability that occurs due to assignable causes in today's manufacturing environments. Control charts are SPCC tools that are mostly used in detecting the variability in processes. One of the earliest SPCC technique developed by Shewhart in (1931) for manufacturing environments involves $\bar{X}$ and R charts.

The $\bar{X}$ chart is used for monitoring the process of location (Mean), and the R chart is used for Variance (dispersion). Later these were extended by Deming in the 1930's. Other control charts for location are the Median and Midrange charts and charts for dispersion include also the S (Standardeviation) and S^2 (variance) charts. The natural variation is always related to the output (products) measurement. It doesn't matter how well the process is designed. Shewhart's technique for monitoring the variations of the process is related to this idea.

Other SPCC methods consist of Cumulative Sum (CUSUM) charts developed in Britain in the 1950's and Exponentially Weighted Moving Average (EWMA) charts (originally proposed by Roberts in 1959) which perform better than Shewhart charts especially in monitoring small process mean and variance changes. All these charts are non-robust charts even a single outlier can degrade their performance. They have zero breakdown point mean that they can't handle even a single outlier.

In general the control charts have three parameters (lines). The central line represents the average level for the in-control process, the other two parameters called upper control line (UCL) and lower control line (LCL). These parameters are chosen and suppose that all the output (products) will lie within these limits as long as the process stay stable or the process remain in control.

1.2. Artificial Neural network (ANN)

In the begining artificial neural networks (ANN) were inspired by biological finding related to the behavior of the brain as a network of units called neurons. An artificial neural network (ANN) is a mathematical model that tries to simulate the structure and/or function aspects of biological neural network. These days the artificial neural network (ANN) are widely used in different fields, like in the field of machine learing , neural networks is used as a major paradigm for Data Mining applications. The modern era of neural network for research was started by the neuro-physiologist McCulloch and the young mathematician Pitt in 1940's. The next major development in neural network technology arrived in 1949 with a book "The organization of Behavior" written by Hebb, D. There are three types of ANN; these are Supervised, Unsupervised and Reinforcement learning.
Usually they are used for any type of network architecture. Supervised neural network training algorithms are used for those cases where researchers require desired outputs with regard to the different inputs. The information is stored in the interconnecting weights and the weights between input vector and output vector are adjusted on the basis of the training algorithm. Different network training algorithms have the ability to determine different types of neural network.

Back-propagation or propagation of error, is a well known supervised training algorithm for teaching artificial neural network how to perform a given task. Supervised ANN is more preferable than unsupervised and reinforcement due to good structure and connection among the input, hidden layers and output with help of weights.

For example in the Multilayer perceptron (MLP) feed forward network, neurons are connected in a series of layers. The Information flow goes in one direction where input is connected with different weights. MLP networks are also known as back-propagation neural networks due to back propagation training algorithm.

1.3. Research Objective

This thesis gives attention to the use of artificial neural networks for monitoring process variability of systems rather than using SPCC techniques.The objective is to develop a network that perform well for detecting any shift the process variability and also be robust for non normal data. To achieve this objective, we will train an artificial neural network (ANN) with the Back propagation technique.

The specific objectives of this research are the following

- To train an ANN for quick detection of out of control signals for small and large shifts in variance.
- To compare the performance of the trained ANN with R charts (for large shift) and EWMA charts (for small shift) in terms of Average run length.
- To test the trained ANN for robustness against non normal data. For this purpose the uniform distribution is used.

2. Statistical Control Charts

A lot of work has been done in the field of statistical quality control to monitor location and dispersion parameters of any process. Statistical control charts like R charts and EWMA charts are used to control the process variability for both large and small shifts respectively.

The R control chart is a statistical data analysis technique for determining if a measurement process has gone out of statistical control. The R chart is sensitive to changes in variation in the measurement process. The R chart is based on the sample range and the average of the subgroup range is used to estimate the process standard deviation and performs well under normality.

Other statistical control charts are the EWMA charts which are also used to monitor the process variability of products. The EWMA charts perform well for detecting also small shifts of 0.5 sigma to 2 sigma in process. In the EWMA charts, the magnitude of weighting factor 'λ' is decided by user to detect how older data points affected the current mean/dispersion values compared to more recent. Generally, it is appropriate to chose small value of lambda (between 0 to 0.2) to detect small shifts and large value (between 0.2 to 0.4) to detect large shifts. The EWMA chart uses information from all sub samples. It detects much smaller process shifts than a normal control charts such as the R chart would do.

2.1. Shewhart R charts

The R chart is used to control the process variability of a system. This chart is mostly used to detect relatively large shifts in the process, i.e. of the order 1.5 sigma or above. The R chart is based on the subsample range. Given a subsample the range R is

$$R = x_{max} - x_{min} \tag{2.1.1}$$

where x_{max} and x_{min} are the maximum and minimum values in the subsample respectively. The sample range is commonly and widely used to estimate the true process standard deviation.

Let $R_1, \ldots, R_n$ be the ranges of n subsamples. Denote by $\bar{R}$ their average, i.e.

$$\bar{R} = \frac{R_1 + \cdots + R_n}{n} \tag{2.1.2}$$

The following control limits of the R chart can be deduced.

$$Upper\ control\ limit = D_4\bar{R} \tag{2.1.3}$$

$$Center\ line = \bar{R} \tag{2.1.4}$$

$$Lower\ control\ limit = D_3\bar{R} \tag{2.1.5}$$

The factors $D_4\ and\ D_3$ are table values which can be found in the R charts literature.

The following figure illustrates the R charts limits.

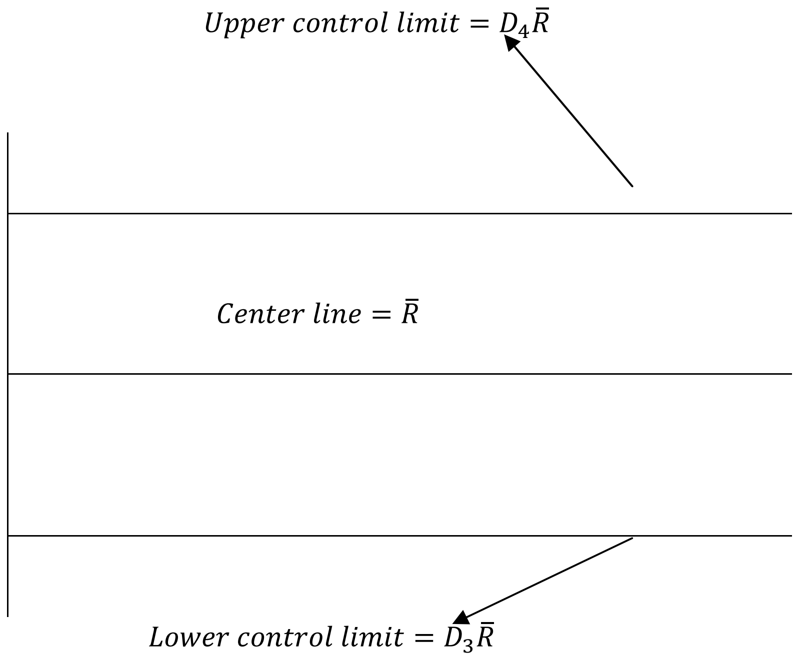


Figure 2.1. Control R charts

If the sample statistic R is plotted outside the upper (2.1.3) and lower (2.1.5) control limits (Montgomery 1991), the process is considered to be out of control.

2.2. EWMA charts

Shewhart control charts like $\bar{X}$ charts or R charts are not able to detect small shifts of location and dispersion parameters of system respectively. The alternative EWMA charts methodology is a more appropriate tool for detecting small shifts of location and dispersion parameters. In general an EWMA statistics $EWMA_i$ is defined as (see Montgomery (2001))

$$EWMA_i = \lambda Q_i + (1-\lambda)EWMA_{i-1} \qquad \text{for } i = 1,2 \ldots n. \tag{2.2.1}$$

where the random variable Q_i of (2.2.1) may be any sample statistic like $x, \bar{x}, R$ or something else but here we use $Q_i = \frac{\sum_{j=1}^{m}(x_{ij}-\mu_0)^2}{m} = \frac{\sum_{j=1}^{m} x_{ij}^2}{m}$, where x_{ij} is subsample value j and sample value at i.

The starting value of $EWMA_o = \sigma^2$ and λ (lying between 0 and 1) is the weighting factor know as the smoothing constant.

The variance of the test statistic defined in (2.2.1) is given (see Montgomery (2001))

$$\sigma^2_{EWMA_i} = \sigma^2_{Q_i}(\frac{\lambda}{2-\lambda})\left[1-(1-\lambda)^{2i}\right] \qquad (2.2.2)$$

where $V(Q_i) = \sigma^2_{Q_i} = \frac{V\left(x_{ij}^2\right)}{m} = 2\sigma^4/m$ is the variance of the random variable Q_i and σ^2 is the variance of x_{ij}. Similarly the mean of the sample statistic defined (2.2.1) is

$$\mu_{EWMA_i} = \sigma^2 \qquad (2.2.3)$$

This means that $\sigma^2_{EWMA_i}$ and μ_{Q_i} are the variance and mean values of the sample statistic $EWMA_i$ respectively.

Hence the k-sigma limit is

$$UCL = \mu_{Q_i} + k\sigma_{Q_i}\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \qquad (2.2.4)$$

which is then

$$UCL = \sigma^2 + k\sigma^2\sqrt{2/m}\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} = \sigma^2(1+k\sqrt{\frac{2\lambda}{m(2-\lambda)}[1-(1-\lambda)^{2i}]}) \qquad (2.2.5)$$

The limits of EWMA chart are thus

$$UCL = \sigma^2\left(1+k\sqrt{\frac{2\lambda}{m(2-\lambda)}[1-(1-\lambda)^{2i}]}\right) \qquad (2.2.6)$$

$$Center\ line = \sigma^2 \qquad (2.2.7)$$

$$LCL = \sigma^2(1-k\sqrt{\frac{2\lambda}{m(2-\lambda)}[1-(1-\lambda)^{2i}]}) \qquad (2.2.8)$$

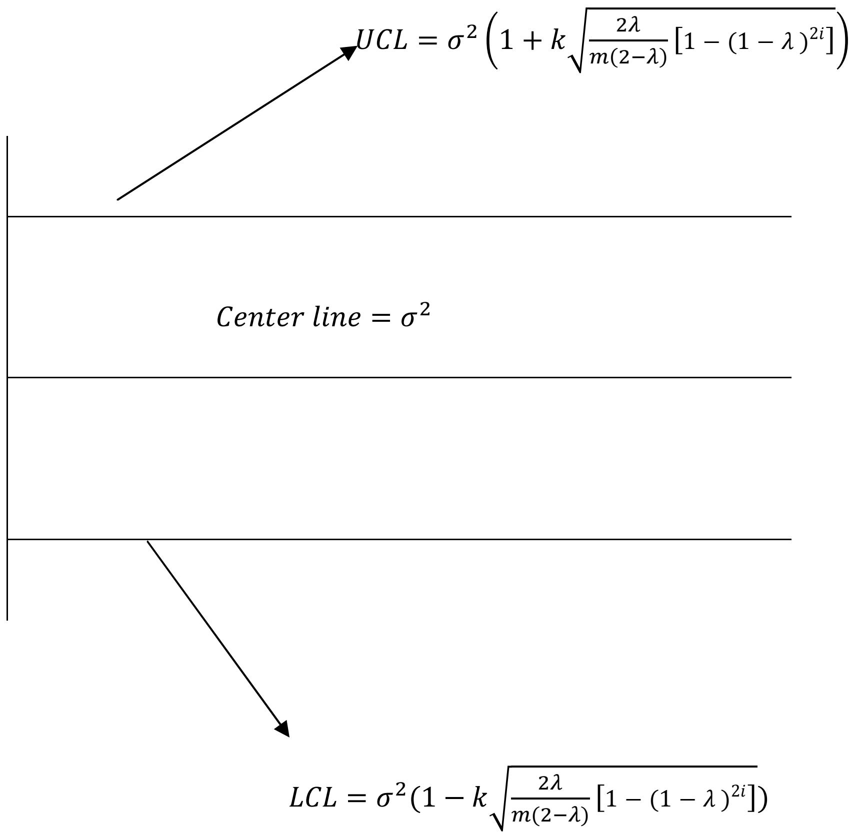


Figure 2.2. Control EWMA charts

If the $EWMA_i$ value is plotted outside the upper (2.2.6) and lower (2.2.8) control limits (Montgomery 1991), the process is considered to be out of control.

In this research the upper and lower values of the R charts as well as for the EWMA charts are chosen by the empirical distribution method. This mean that after generating a large data set, we find the upper and lower critical values as particular percentiles of the empirical distribution of the chart statistic for an in-control process.

3. Artificial neural network (ANN) approach for process variance shift detection

Here the supervised technique of ANN is used to detect the process variability for both large and small shifts. For our study purpose supervised technique of ANN with back propagation network is best because according to different inputs we require different desired outputs. Here the main problem is to find out the numbers of hidden layers and nodes for each layer for the minimum error and the choice of learning rate that gives optimal weights for network for getting the desired ouput. For the selection of hidden layers and nodes we adapt the rule as described by Chang and Ho (1999) and thus select an ANN with topology [5-12-12-1], meaning 5 input values, and two hidden layers each consisting of 12 nodes and one output.

For training a best network the following procedure is used

- Presenting the pair of Input/output vector.
- Determining the error between actual and desired output.
- Propagating the error back to the network to adjust weights and thereby reducing the error.
- For getting the optimal weights for the best network the choice of adaptive learning rate.

We will train the MLP with Back Propagation with five normal distributions consisting of a total of 600 samples for the whole procedure. Each distribution, with a different variance, will generate 120 sub-samples. Here $x_{jk\rho} \sim N(0, (\rho\sigma_0)^2)$, k =1,...120 is the raw data where j=1,…,5 the index of the observation that within a sample and $\rho = 1, \dots, 5$ represent the different shifts for the five distributions. Here $\sigma_0 = 1$ throughout the research. The data for each group are transformed by $z_{jk\rho} = |x_{jk\rho} - \bar{x}_{k\rho}|$ and these values are inputs to the ANN. Due to its robustness in statistical tests for equality of variances, where $\bar{x}_{k\rho} = \frac{1}{5}\sum_{j=1}^{5} x_{jk\rho}$. The samples from the different distributions are shown in table 3.1.

Table 3.1. Input data for training the ANN

Group	No. Of samples	Distribution $\boldsymbol{N(0, (\rho\sigma_0)^2)}$	shift $\boldsymbol{\rho = \sigma/\sigma_0}$
A	120	$\boldsymbol{N(0, 1^2)}$	1
B	120	$\boldsymbol{N(0, 2^2)}$	2
C	120	$\boldsymbol{N(0, 3^2)}$	3
D	120	$\boldsymbol{N(0, 4^2)}$	4
E	120	$\boldsymbol{N(0, 5^2)}$	5

3.1. Artificial neural network (ANN) Algorithm with Multilayer Perceptron (MLP)

The following figure shows the artificial neural network that deals with five input values, z_{jk}, j=1,...5, at a time with two hidden layers, each with 12 nodes, produce one single output. Here the index ρ has been omilted.

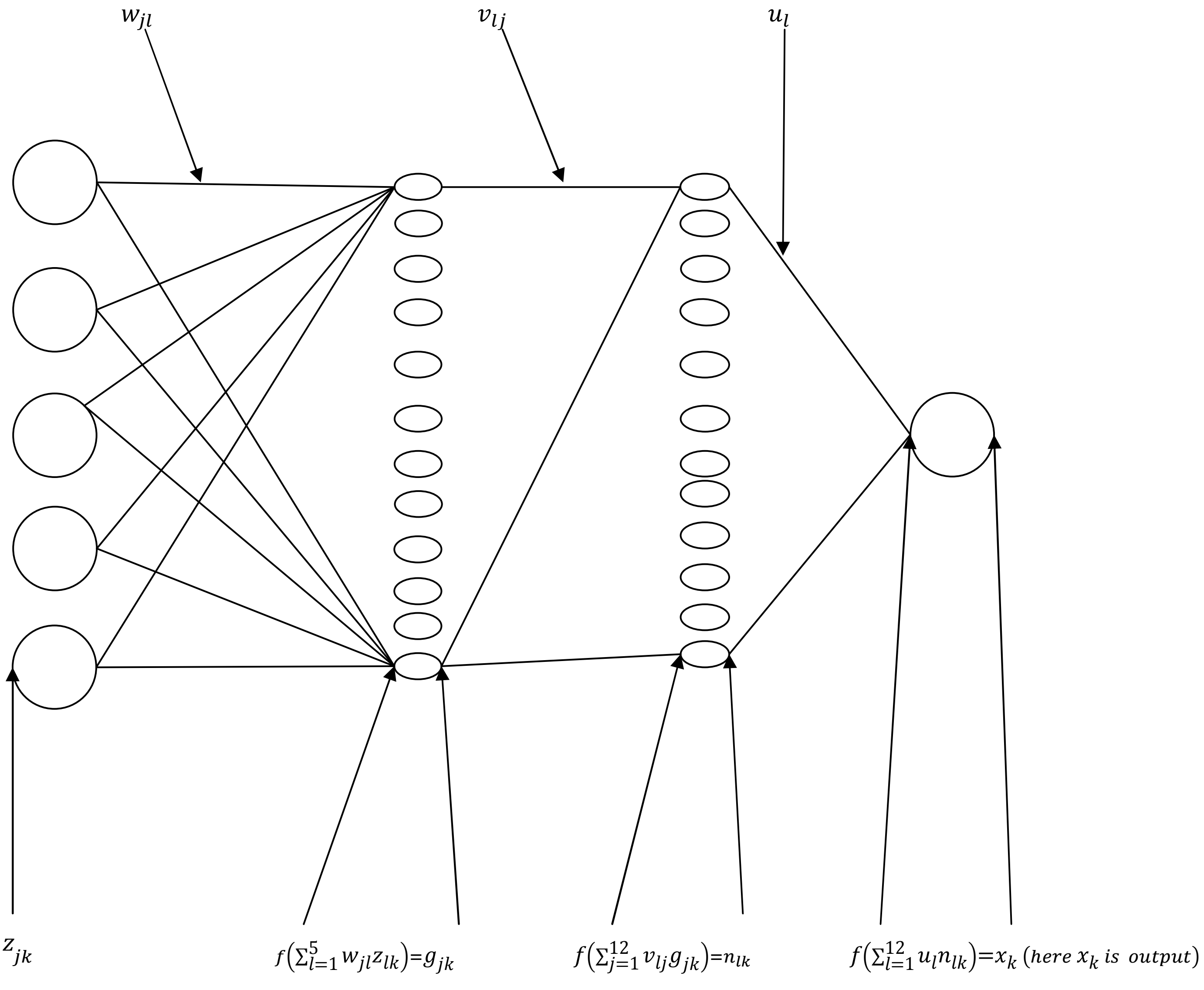


Figure 3.1. ANN Multilayer Perceptron structure

Every node generate an output that is the function $f(x)$ of the linear combination of the output from previous nodes. The training algorithm for MLP algorithm requires a differentiable, continuous nonlinear activation function $f(x)$. The most used function for Back propagation is the logistic function,

$$f(x) = \frac{1}{(1+e^{-x})} \tag{3.3.1}$$

which is differentiable with

$$f(x)'=f(x)(1-f(x)) \tag{3.3.2}$$

The final output x_k is the function of $f\left(\sum_{l=1}^{12} u_l n_{lk}\right)$, which is the linear combination, with weights u_l of outputs n_{lk} from nodes in the last hidden layer. The outputs n_{lk} and g_{jk} are generated in a similar way.

The ANN receive z_{jk} values as input and produce the different outputs x_k for different shifts. For the different shifts the ANN will produce the different outputs.

The expected desired outputs, denoted by $\boldsymbol{x^d}$, from ANN for different shifts are shown in table 3.2.

Table 3.2. Desired output from ANN

Group	Desired Outputs=$\boldsymbol{x^d}$	Interpretation
A	0.050	$\boldsymbol{\rho = 1}$ no shift
B	0.275	$\boldsymbol{\rho = 2}$ shift
C	0.500	$\boldsymbol{\rho = 3}$ shift
D	0.725	$\boldsymbol{\rho = 4}$ shift
E	0.950	$\boldsymbol{\rho = 5}$ shift

The desired outputs are decided on theoretical basis according to the different shifts. Outputs for shifts one to five are observed and it was decided that the ANN should produce desired outputs defined in (table 3.2) for the different shifts.

If the proposed ANN is not able to produce the desired outputs for all shifts, it is necessary to train the ANN by the back propagation method to obtain optimal weights.

3.2. Selection Criteria of optimum weights

The optimal weights for the ANN strcuture is obtained by the back propagation method. Initially the weights of the network are randomly assigned values between -1 to 1. The back propagation algorithm is then used to compute the necessary correction. The ANN algorithm can be decomposed in the following four steps.

- Feed –forward computation
- Back propagation to the output layer
- Back propagation to the hidden layer

- Weights updated

The performance criteria of the training network for getting optimum weights is totally depending on the magnitude of the loss function,

Loss Function= $E = \frac{1}{2}\sum_{A,B,C,D,E}(desired\ output - actual\ output)^2$ (3.2.1)

where $A, B, C, D\ and\ E$ are the representatives of different groups of distribution as shown in table 3.1. The aim is to get the optimum weights for whole network by minimizing the loss function, which is the sum of squared errors. The ANN algorithm is stopped when the value of the loss function has become sufficiently small.

3.3. Mathematical derivations of optimal weights with the back propagation method

Let x_k be the output from the kth input $z_{1k}, z_{2k}, z_{3k}, z_{4k}, z_{5k}$. Based on the output n_{lk} from the lth node in the second hidden layer and weights u_l the following is the expression of the output:

$$x_k = f(\sum_{l=1}^{12} u_l n_{lk}) = \frac{1}{(1+e^{-\sum_{l=1}^{12} u_l n_{lk}})} \quad (3.3.1)$$

Let x^d denote the desired output, then the loss function is

$$E = \frac{1}{2}\sum_{k=1}^{120}(x^d - x_k)^2 = \frac{1}{2}\sum_{k=1}^{120} e_k^2 \quad (3.3.2)$$

where $e_k = x^d - x_k$

The optimal weights of the ANN algorithm for the whole network will be updated using the gradient descent rule.

Let T denote the whole set of weights w, v, u. Given an initial set of weights T, we find a better set of weights T by updating the ANN using the Gradient descent rule, that is

$$T(t+1) = T(t) - h(dE/dT) \quad (3.3.3)$$

where h =learning rate.

Gradient descent rule for optimal weights between output and last hidden layers is

$$u_l(t+1) = u_l(t) - h(dE/du_l) \quad (3.3.4)$$

The differential of E in equation(3.3.2) w.r.t the weight u_l is

$$\frac{dE}{du_l} = -\sum_{k=1}^{120} e_k (dx_k/du_l)$$

$$= -\sum_{k=1}^{120} e_k f' \left(\sum_{l=1}^{12} u_l n_{lk}\right) \frac{d}{du_l} \left(\sum_{l=1}^{12} u_l n_{lk}\right)$$

$$= -\sum_{k=1}^{120} e_k f \left(\sum_{l=1}^{12} u_l n_{lk}\right)\left(1 - f\left(\sum_{l=1}^{12} u_l n_{lk}\right)\right) n_{lk}$$

$$\frac{dE}{du_l} = -\sum_{k=1}^{120} e_k x_k (1 - x_k) n_{lk} \qquad (3.3.5)$$

Now for updating the weight u_l, put (3.3.5) in (3.3.4) to obtain

$$u_l(t+1) = u_l(t) + h \sum_{k=1}^{120} e^k \, x_k (1 - x_k) \, n_{lk} \qquad (3.3.6)$$

In a similar way we can find the optimal weights for second and first hidden layers. The corresponding updating formulas are :

$$v_{ab}(t+1) = v_{ab}(t) + h \sum_{k=1}^{120} e_k \, x_k (1 - x_k) \, u_a n_{ak} (1 - n_{ak}) \, g_{bk}$$

and

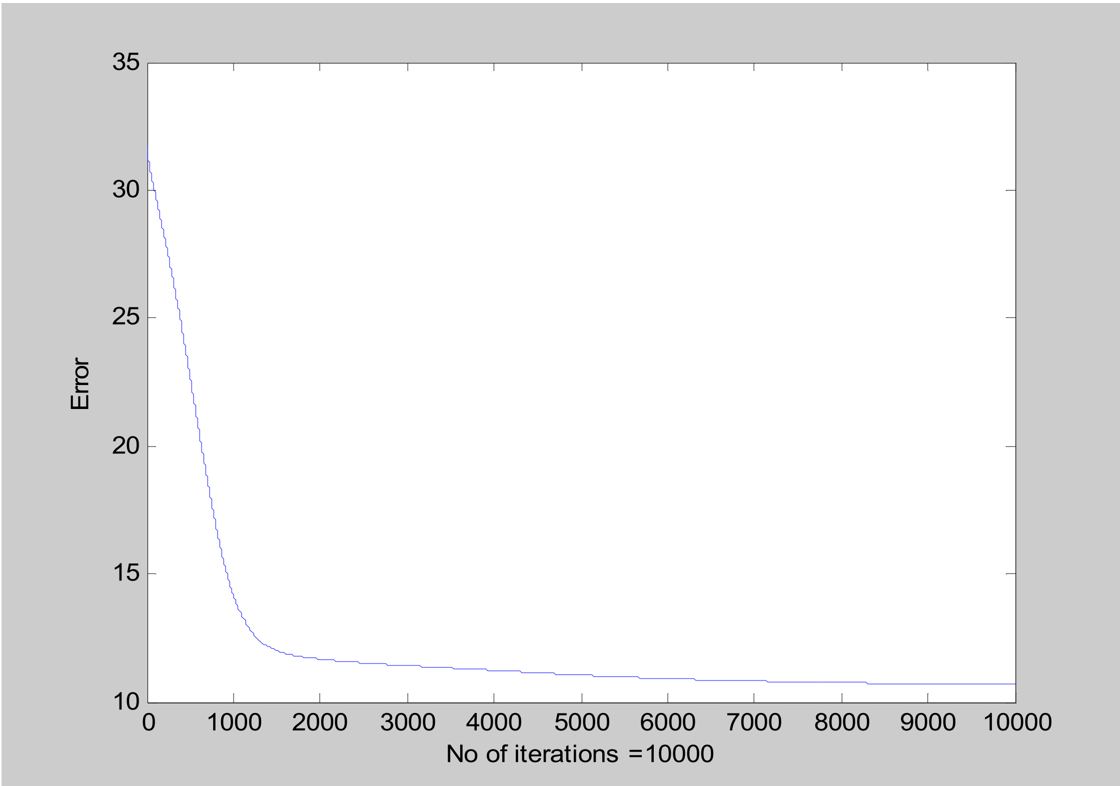
$$w_{ac}(t+1) = w_{ac}(t) + h \sum_{k=1}^{120} e_k \, x_k (1 - x_k) g_{ak} (1 - g_{ak}) z_{ck} \sum_{l=1}^{12} u_l \, n_{lk} (1 - n_{lk}) v_{la}$$

4. **Simulated results with trained ANN for different shifts**

The numbers of inputs, hidden layers, nodes for each hidden layer and outputs are fixed as described previously according to our inputs and desired outputs requirement and learning rate is $h = 0.001$, because small learning rate quickly finds out the optimal weights with minimum error. We trained the proposed ANN for optimal weights and found the minimum error in (3.2.1) for optimal weights for proposed network to be,

$$E = 10.66$$

Figure 4.1. Error behavior with 10000 iteration



We tested our trained network with the random inputs data simulated from normal distributions $x_{jk\rho} \sim N(0, (\rho\sigma_0)^2)$ where $\rho = 1,2, \dots ,5$ with 10000 sub-samples of five for each distribution. The outputs are shown in table 4.1.

Table 4.1. Percentage of ANN outputs falling into specified output range between 0 and 1 for different shifts. Results are based on 10000 inputs set, each of five values, for each distribution.

Output range	Input data $\rho = 1$ vs target=0.05	Input data $\rho = 2$ vs target=0.275	Input data $\rho = 3$ vs target=0.5	Input data $\rho = 4$ vs target=0.725	Input data $\rho = 5$ vs target=0.95
0.00-0.05	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0%**
0.05-0.10	17.04%	1.53%	0.35%	0.12%	0%
0.10-0.15	33.47%	5.42%	1.40%	0.51%	0.24%
0.15-0.20	22.80%	7.96%	2.17%	0.80%	0.36%
0.20-0.25	13.68%	9.12%	2.88%	1.31%	0.59%
0.25-0.30	6.96%	10.32%	3.68%	1.36%	0.61%
0.30-0.35	3.55%	10.41%	4.51%	1.84%	0.85%
0.35-0.40	1.57%	9.87%	5.60%	2.48%	1.17%
0.40-0.45	0.61%	10.10%	6.42%	3.32%	1.65%
0.45-0.50	0.23%	8.98%	7.20%	4.40%	2.10%
0.50-0.55	0.07%	8.12%	8.69%	5.25%	2.98%
0.55-0.60	0.02%	6.50%	9.76%	6.70%	4.00%
0.60-0.65	0.00%	5.39%	10.72%	8.52%	5.51%
0.65-0.70	0.00%	3.61%	11.13%	11.15%	7.91%
0.70-0.75	0.00%	1.93%	10.68%	13.29%	11.33%
0.75-0.80	0.00%	1%	8.60%	14.55%	15.85%
0.80-0.85	0.00%	0.16%	4.75%	14.88%	20.39%
0.85-0.90	0.00%	0.01%	1.37%	8.35%	19.87%
0.90-0.95	0.00%	0%	0.09%	1.16%	4.40%
0.95-1.00	0%	0%	0.00%	0.01%	0.15%

As we observe from the table 4.1, the ANN can't produce the desired outputs for the each shift due to error. But the outputs for each shift scatter around the target values as shown in table 4.1. The outputs of ANN for $\rho = 1$ are not exactly as the desired output should be '0.05' but they are scattering around the desired output and its neighborhood. Similarly when $\rho = 2$ the desired output is '0.25' but the outputs are scattering around the target value and its neighborhood.

According to the percentage of outputs of ANN a cut point will be selected as critical value for out-off control signal. This means that if the outputs of ANN greater than a cut point, the process is considered to be out of control.

A plot of th percentage data (as shown in table 4.1) is shown in figure 4.2.

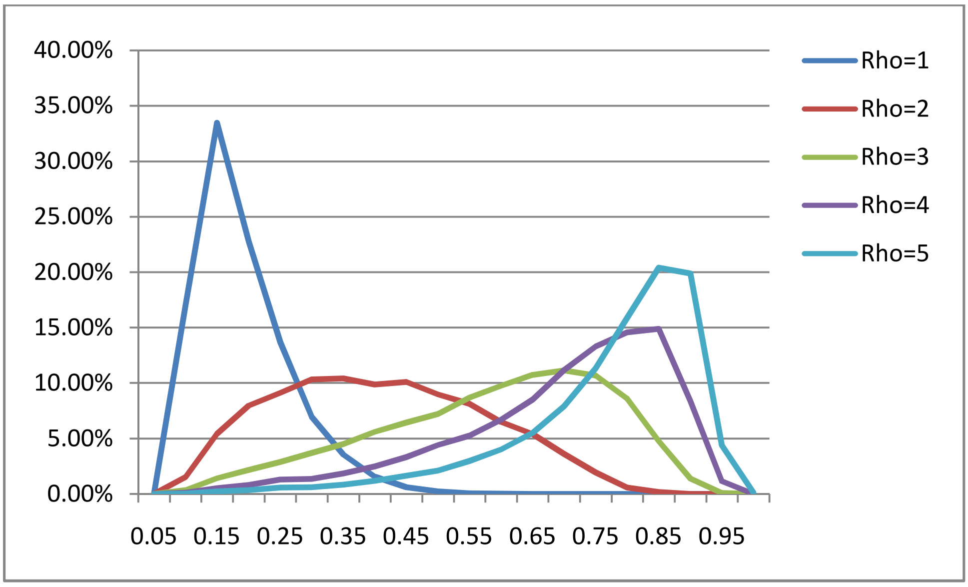


Figure 4.2. Behavior of percentage of different shifts

Since the outputs of all shifts of ANN scatter around the desired values for a particular process, a good estimate of variance change magnitude cannot be obtained. For desired outputs for all shifts our error in (3.2.1) should be zero.

5. Performance analysis of the proposed ANN

Monte Carlo simulation is used to check the performance of the proposed ANN methodology. The performance of the ANN methodology is compared with statistical control charts. The critical values, cut-off-point for ANN and upper and lower limits for R and EWMA charts (as described in sections 2.1 and 2.2 for $\rho\sigma_o = 1$) are chosen by empirical distribution method by assuming 5% error probability. An average run length (ARL) criteria is used to check the performance of ANN and statistical control charts for variance change detection.

The ARL of ANN will be compared with R and EWMA charts for normally distributed samples. To test the ability that training network is robust for any kind of data for variance change detection, uniform distribution is used. For comparison, here we replicated 10,000 times data from $x_{jk\rho} \sim N(0, (\rho\sigma_0)^2)$ where $\rho = 1,2, \dots ,5$, for ANN technique, R charts and EWMA charts and for uniform distribution with same variances as for the normal distribution and calculated ARL.

5.1. Selection criteria of critical values by Empirical distribution method

The empirical distribution method is used to find out the cut-off-point for ANN and upper and lower limits for R and EWMA charts. By Monte Carlo simulation we replicated 10000 values for ANN, R and EWMA charts and sorted them and found the 95 percentile value for ANN as cut-off-point, 0.025 percentile and 0.975 percentile values as lower and upper limits for R and EWMA charts respectively, assuming the 5% error probability. The cut-off-point of the ANN was found to be is '0.3134' and the upper and lower values of R and EWMA charts are as follows.

For R charts

$$\left.\begin{array}{c} upper\ control\ limit = 4.20 \\ center\ line = 1 \\ lower\ control\ limit = 0.85 \end{array}\right\} \tag{5.1.1}$$

For EWMA charts

$$\left.\begin{array}{c} upper\ control\ limit = 1.31 \\ center\ line = 1 \\ lower\ control\ limit = 0.7 \end{array}\right\} \tag{5.1.2}$$

5.2. Average run length (ARL)

There are two ways to analyse the average run length (ARL)

- Fix the in-control average run length and compare out-of-control average run lengths. Methods with smaller out-of-control average run length are better than others.
- Fix the out-of-control average run length and compare in-control average run lengths. Methods with larger in-control average run length are better than others.

  The expression for in-control ARL

$$ARL_{in-control} = \frac{1}{p(Type\ \mathrm{I}\ error)} = 1/\alpha \tag{5.2.1}$$

  Here $\alpha = \text{error probability}$

The expression for out-of-control ARL

$$ARL_{out-control} = \frac{1}{1-p(Type\ \mathrm{II}\ error)} = 1/(1-\beta) \tag{5.2.2}$$

Since ANN, R and EWMA charts have the same in-control ARL as

$$ARL_{in-control} = \frac{1}{p(Type\ \mathrm{I}\ error)} = \frac{1}{\alpha} = \frac{1}{0.05} = 20. \tag{5.2.3}$$

the out-of-control ARL is the best measure of relative chart effectiveness. The out-of-control ARL directly determines the average number of sample taken to signal after a shift occurs in process. Therefore, it is desirable to have lower out-of-control ARL.

5.3. Comparison of ANN and Statistical control charts(R & EWMA) in term of ARL

The comparison of ANN and Statistical control charts is based on the magnitude of average run length (ARL). As described previously we will compare out-of-control ARL. Between ANN and SPCC, which posses less ARL is perform well as compare to each other. For comparison ARL are given in table 5.4.

Table 5.4. Comparison of ARL among ANN, R chart and EWMA charts

$\rho = \frac{\sigma}{\sigma_o}$	ANN output >0.3134	Standard R charts	EWMA charts
1.00	20.6815	22.9735	21.9435
1.20	6.8068	8.8660	9.7386
1.40	3.5337	4.7398	4.8552
1.60	2.4082	2.8056	3.1383
1.80	1.8748	2.1510	2.3665
2.00	1.5768	1.7555	2.2840
3.00	1.1326	1.1593	1.8619
4.00	1.0438	1.0545	1.2128
5.00	1.0182	1.0254	1.0231

ARL values of ANN, R and EWMA charts are in columns two, three and four respectively. The ARL values of ANN as compare to ARL values of R and EWMA charts are less for all different shifts .This mean ANN play significant role to detect variance change shifts as compare to R and EWMA charts techniques. The proposed methodology of ANN is performing well in the term of ARL to detect variance change for both small and large shifts.

5.4. Performance of trained ANN compared with uniformly distributed data

As mention earlier the trained ANN is not only suitable for normal data but is also performing well for non normal data. Here robustness of the trained ANN is tested with the uniform distribution. When the trained ANN deals with uniformly distributed data, it behaves differently, the cut-off-point as a critical value will be changed. The cut-off-point was selected by the empirical method, and is ‘0.2892’ by assuming 5% error probability. The ARL values are given in table 5.4 for comparison for both ANN with normal data and ANN with uniformly distributed (ANNU) data.

Table 5.4. Comparison of ARL between ANNU and ANN .

$\rho = \frac{\sigma}{\sigma_o}$	ANNU output > 0.2892	ANN output > 0.3134
1.00	20.3855	20.6815
1.20	4.9239	6.8068
1.40	2.5379	3.5337
1.60	1.7668	2.4082
1.80	1.4507	1.8748
2.00	1.2951	1.5768
3.00	1.0607	1.1326
4.00	1.0206	1.0438
5.00	1.0072	1.0182

The ARL values for all shifts when ANN is dealing with non-normal and normal data are given in colums two and three respectively. The ALR magnitude of ANNU for all shifts are small as compare to ANN ARL magnitude. Therefore proposed methodology of ANN is also performing well to detect process variability for both small and large shifts for non normal data.

6. Conclusions and Discussion

The proposed methodology of ANN performs well for process variance change detection for both small and large shifts. Simulated results show that proposed artificial neural network approach is more efficient than traditional statistical control charts for process variance change detection in term of ARL. In addition, the ANN methodology is also performing well when data are generated from non-normal distributions.

In this research, the proposed ANN is only trained for normal data, for more efficient results for variance change detection ANN can be trained for any kind of data like uniform distribution and may be used for any other data. It works efficient when all normal distributions have means zero. However we can train an ANN that can handle both shifts in mean and variance level, and which at the same time is robust for non normal data. This research will however be presented else where.

References

Benneyan, J. C. (1998). Use and interpretation of statistical quality control charts. *International Journal of Quality in Health Care*, Vol 10, 69-73.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Singapore: Springer.

Box, G. E. P. and Kramer, T. (1992). Statistical process monitoring and feedback adjustment. *Technometrics*, Vol 34, 251-285.

Box, G. E. P. , Coleman, D. E. and Baxley, R. V. (1997). A comparison of statistical process control and engineering process control. *Journal of Quality Technology*, Vol 29, 128-130.

Chang, S. I. and Ho, E. S. (1999). A two stage neural network approach for process variance change detection and classification. *International Journal of Production Research*, Vol 37, 1581-1599.

Cheng, T. C. and Gan, F. F. (1995a). A multi-layer neural network model for detecting changes in the process mean. *Computers in Industrial Engineering*, Vol 28, 109-119.

Cheng, T. C. and Gan, F. F. (1995b). A neural approach for the analysis of control chart pattern. *International Journal of Production Research*, Vol 35, 667-697.

Guo, Y. and Dooly, K. J. (1992a). The application of neural networks to a diagnostic problem in quality control. Proceeding of ASME-WAM: Monitoring and Control for Manufacturing Processes, 111-122.

Guo, Y. and Dooly, K. J. (1992b). Identification of change structure in statistics process control. *International Journal of Production Research,* Vol 30, 1655-1669.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Montgomery, D. C. (1997*). Introduction to Statistical Quality Control*, 3rd Edition, New York: Wiley.

Pagh, G. A. (1991a). Synthetic neural networks for Process control. *Computer and Industrial Engineering*, Vol 17, 24-26.

Pagh, G. A. (1991b). A comparison of neural network to SPC charts. *Computer and Industrial Engineering*, Vol 21, 253-255.

Riaz, M. and Sagir, A. (2007). Monitoring Process Variability Using Gini's Mean Difference. *Quality Technology and Quantitative Management*, Vol 4, 439-454.

Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, Vol 1.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: Van Nostrand.

Smith, A. E. (1993). X-bar and R control chart interpretation using neural computing. *International Journal of Production Research*, Vol 32, 309-320.

Smith, A. E. and Dagli, C. H. (1991). Controlling industrial processes through supervised, feedforward neural networks. *Computer and Industrial Engineering*, Vol 21, 247-251.

Smith, A. E. and Yazici, H. (1992). An intelligent composite system for statistical process control. *Engineering Application of Artificial Intelligence*, Vol 5, 519-526.