An application of artificial neural network to maintenance management

V. O. Oladokun*

Lecturer, Dep. of Industrial and Production Engineering, University of Ibadan, Ibadan, Nigeria

O. E. Charles-Owaba

Associate Professor, Dep. of Industrial and Production Engineering, University of Ibadan, Ibadan, Nigeria

C. S. Nwaouzru

Dep. of Industrial and Production Engineering, University of Ibadan, Ibadan, Nigeria

Abstract

This study shows the usefulness of Artificial Neural Network (ANN) in maintenance planning and management. An ANN model based on the multi-layer perceptron having three hidden layers and four processing elements per layer was built to predict the expected downtime resulting from a breakdown or a maintenance activity. The model achieved an accuracy of over 70% in predicting the expected downtime.

Keywords: Maintenance management; Downtime; Artificial Neural Network; Multi-layer perceptron

1. Introduction

Maintenance management system is crucial to the overall performance of any production system. Productivity depends on the effectiveness of the maintenance system, while effective and efficient production planning depends on the ability of the planners to adequately predict some crucial maintenance systems' parameters. Parameters such as Mean Time To Failure (MTTF), Mean Time Between Failures (MTBF), mean repair time and Downtime (DT) are required for adequate characterization of most maintenance environments. Having an adequate knowledge of these maintenance parameters for a given production system, we will enhance other functions like the scheduling and planning functions thereby improving overall system performance.

The literature is replete with the application of Operations Research tools to various maintenance related problems. For example, Bobos and Protonotarios [6] applied the Markov model to the area of equipment maintenance and replacement problem. Aloba et al. [2] used simulation for the planning of preventive maintenance activities. In particular,

mathematical models have been used to study the downtime characteristics associated with various maintenance environments [7,11]. In this study, the concept of artificial neural network is applied to the problem of predicting the downtime associated with a facility breakdown. The downtime cost has been identified as a major maintenance cost component [4].

2. Problem description

For a typical manufacturing system the duration of each downtime (DT) resulting from a breakdown or a maintenance activity is a random variable, where *Downtime = machine breakdown period + maintenance repair tim*e. Since DT is a function of some two important maintenance parameters and because DT affects the overall available production period, being able to predict the expected downtime from a given breakdown and maintenance activity will therefore be useful for both short term and long term production planning. In other words, given a machine breakdown it will be useful to say whether the result-

---

* Corresponding author. E-mail:vo.oladokun@mail.ui.edu.ng

ing DT will be short, medium or long term by extrapolating from historical data of breakdown types and the associated DTs. If it is envisaged that the resultant DT will be a short-term one then production plan may not be altered, job orders may not be canceled, overtime may not be necessary, however if it is envisaged that the DT will be a prolonged one then production management can be advised on the appropriate steps to be taken in order to mitigate the effect of such DT. Note that classification of DT into short, medium, or long will depend on the specific production environment.

Because a production system tends to be subject to random failures arising from continuous use and external factors, most of the crucial maintenance related parameters tend to be random or stochastic in nature. Conceptually, the duration of any downtime DT is expected to be a function of some number of maintenance and production factors (parameters) prevailing prior to the break down. It is obvious that this function will not be simple or linear in nature. Mathematical models reported in the literature are usually based on some simplifying assumptions. For instance, Brouwers [7] derived mathematical expressions for the downtime of an equipment assuming exponential failure and repair distributions for the equipment.

Also, a MTTF much larger than the mean time for repair was also assumed. Similarly, Kiureghian et al. [11] in order to derive closed-form expressions for the mean duration of downtime assumed that component failures are homogeneous Poisson events in time and that repair durations are exponentially distributed. Unfortunately, these assumptions limit the applications of the models in many real life problems hence the need for other approaches.

A practical approach to this type of problem is to apply conventional regression analysis in which historical data are best fitted to some functions. The result is an equation in which each of the inputs x_j is multiplied by a weight w_j; the sum of all such products and a constant θ gives an estimate of the output as follows:

$$y = \sum_j w_j x_j + \theta$$

The drawback here is the difficulty of selecting an appropriate function capable of capturing all forms of data relationships as well as automatically modifying output in a case of additional information. An Artificial Neural Network (ANN) which imitates the hu-

man brain in problem solving, is a more general approach that has these desirable attributes [12,13]. A general overview of the ANN is contained in the Appendix I.

3. Methodology

The firm in the case study is a high volume commercial packaging firm based in Lagos city, Nigeria. The firm prints on packaging cartons, packs and papers for clients. It operates a set of coding machines which are now fairly old and thus require substantial maintenance attention. The maintenance of these willet coding machines is carried out in-house by the maintenance department. The data used in this study were collected from the maintenance department activities record.

The firm was observed for a number of days, experienced production and maintenance staff were interviewed for historical details and operating records vetted. Having thus grasped the total picture of the maintenance environment, the required input variables, which influence the duration of repairs of any particular breakdown, were identified.

An in-depth analysis of breakdown records of the input variables (factors) was carried out to determine their likely influence on the output variable. These input variables were grouped into two categories; namely machine and product. These were, in turn, classified into minor, major and catastrophic faults. The choice of these faults is based on the rationale that any downtime is preceded by some series of faults either in the machine, or the product so that the nature of the final breakdown may be stored in the pattern of the preceding faults. In other words, each fault pattern or a combination can be associated with a given length of downtime. The idea is to use these classifications to predict the type of breakdown they will cause, since the nature of a breakdown is reflected in the time taken to correct it.

After the classification, the data were collected and transformed into a form suitable for neural network coding and then the general topology of the neural network for predicting the DT was designed.

3.1. Neural network training performance

The training performance is then evaluated using the following performance measures, namely the Mean Square Error (MSE) and Correlation Coefficient (R):

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}(d_{ij}-y_{ij})^2}{NP},\qquad(1)$$

where,

P Number of output of processing element.

N Number of exemplars in the data set.

y_{ij} Network output for exemplars i at processing element j.

d_{ij} Desired output for exemplars i at processing element j.

$$R = \frac{\dfrac{\sum(x_i - x_{mean})(d_i - d_{mean})}{N}}{[(\dfrac{\sum(d_i - d_{mean})^2}{N})(\dfrac{\sum(x_i - x_{mean})^2}{N})]^{0.5}}.\qquad(2)$$

Equation (1) indicates the degree of fit while Equation (2) gives degree of relationship or the direction of movement of the two sets of data, which MSE fails to show.

The last stage is to run the experiment. This and the earlier stated stages were judiciously followed as presented in the following section.

3.2. The data set design

The data set consists of input factors/variables and an output variable. The input variables/factors are production and operational factors as well as maintenance parameters that influence the duration of repair of a particular breakdown. The output variable is the downtime after a given breakdown. The downtime is however classified into three classes of short, medium and long time for effective modeling.

3.2.1. Data collection

Proper data collection plan was done so that:

1) Data collected will be sufficient.

2) The data is as free as possible from observation noise.

The data covering three months of operation were collected from records and discussions with appropriate personnel of the case company based in Lagos.

3.2.2. Machine faults

- *Minor faults*. These are symptoms that when noticed can still be tolerated over a long time though if not even attended to over a tolerable period could cause machine breakdown.

- *Major faults*. These are symptoms that when noticed should be attended to immediately or else they could lead to a major break down of the machine.

- *Catastrophic faults*. These are symptoms, which automatically lead stoppage of the machine and require immediate maintenance repair.

Hence the first three variables are minor machine faults, major machine faults, and catastrophic machine faults. See Table 1 for some common symptoms classification.

3.2.3. Products faults

These are observable faults in the output indicating some problems in the machine. They are also grouped into three categories based on experiences of the operators.

- *Minor Product fault.* These are faults that may not require immediate stoppage. It implies that machine can go ahead and finish a set task or target while fault can be attended to while production is running.

- *Major Product fault.* These are product faults that may necessitate the stoppage of machine so that maintenance repair can be carried out.

- *Catastrophic Product fault.* These types of faults are those that when they occur causes immediate stoppage of machine and also production.

Hence the second sets of variables are minor product faults, major product faults and catastrophic product faults (Table 1).

3.2.4. Data transformation and domain for input data

The collected data has to be expressed (transformed) into a format that optimizes the performance of the neural network. Hence, the domain of the input variables that is used in this study is as shown in Table 2.

3.2.5. Output variable

The output variable represents the downtime (DT) as follows:

DT = machine breakdown period + maintenance repair time

The output variable represents the duration of downtime, classified into three groups based on likely impact on production operation. The three classes of output variables are short term, medium term and long term.

1) *Short term:* A downtime of less than 30 minutes. This length of downtime may not affect production schedule seriously

2) *Medium term:* A downtime between 31 minutes to 120 minutes. This duration may affect production schedule and hence leads to planning of overtime so as to meet target.

3) *Long term:* Two hours and above is considered serious enough to affect production schedule harshly and likely leads to production target to been missed.

Note that classification of DT into short, medium, or long will depend on the specific production environment.

3.2.6. Transformation and domain for output variable

The output variable domain chosen for the three variable classes is shown in Table 3.

Note that classification of DT into short, medium, or long will depend on the specific production environment.

3.3. Neural network building (topology)

After the data has been transformed and the method of training has been chosen, it is then necessary to determine the topology of the neural network. The network topology describes the arrangement of the neural network. The network topologies available for the neural network are numerous, each with its inherent advantages and disadvantages, for example some networks trade off speed for accuracy, some are capable of handling static variables and they are not continuous. Hence, in order to arrive at an appropriate network topology, various topologies such as multi-layer perceptron, recurrent network, and time-lagged recurrent network were considered. Due to the nature of the data available for this study, which is static, the multi-layer perceptron was selected.

3.3.1. Multi-layer perceptron

Multi-layer perceptrons (MLPs) feed-forward networks are typically trained with static data. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input/output map. The key disadvantages are that they train slowly, and require lots of training data, typically three times more training samples than network weights [1].

3.3.2. The network layers and processing elements design

The next step in the building of the neural network model is the determination of the number of processing elements and hidden layers in the network. In the selection of number of processing elements to suit the network a tradeoff has to be made. A large number of processing elements mean large number of weights, though this can give the network the possibility of fitting very complex discriminating functions, it has been seen that too many weights produce poor generalizations. On the other hand very small number of processing elements reduces the discriminating power of a network. Hence the PEs was varied in the study from 1 to 5 nodes, to get the most suitable network. In the running of the experiment observations were made on the behavior of the learning curve. The number of PEs that gives the minimum value Mean Square Error (MSE) for the learning curve is used.

In determining the number of hidden layers to be used, there are two methods in the selection of network sizes: one can begin with a small network and then increase its size (growing method); the other method is to begin with a complex network and then reduce its size by removing not so important components (pruning method). The growing method was used in the building of the neural network model. Hence, the experimentation involves starting with no hidden layers and then gradually increasing them.

3.4. Experimentation

A student version of commercial ANN modeling software was used. Various network models were attempted with variations in the number of hidden layers from 0 to 5 and also processing elements from

Table1. Machine and product faults.

Minor Machine Faults	Major Machine Faults	Catastrophic Machine Faults
Mixer tank low, ink low solvent tank low, fluctuation in sequence of labeling, ink low, viscosity variation.	Drop in jet pressure, unusual noise, reduced speed rotation, high or low viscosity, bad keypad, persistent charge error.	Power fluctuation, fan failed, drop in engine running pressure, not jetting, lid/hood problem, fluctuating phase angle.
Minor Product Faults	**Major Product Faults**	**Catastrophic Product Faults**
Marginal break in print characters, poor print, poor modulation.	Persistent charge error, wrong color print distortion, clipping of code.	Gutter fault, irregular print, no print.

Table 2. Data transformation and domain for input data.

	Input Variable	Domain	Code
1	Major machine faults Minor machine faults Catastrophic machine faults	Present Not present	1 0
2	Major product faults Minor product faults Catastrophic product faults	Present Not present	1 0

Table 3. The output variable domain.

Output Variable	Domain
Short term	1
Medium term	2
Long term	3

Table 4. Confusion matrix.

Output / Desired	Short Time	Medium Time	Long Time
Short time	13	4	0
Medium time	4	7	0
Long time	0	1	1

Table 5. performance measures.

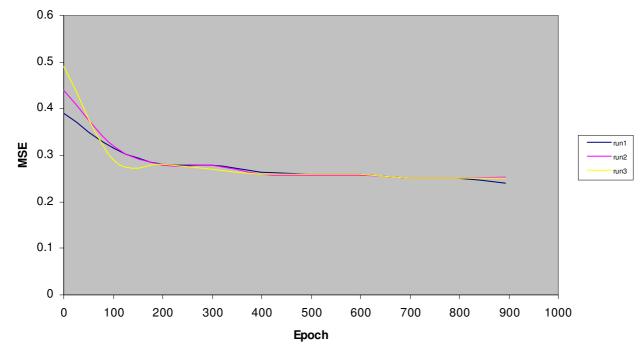| Performance	Short Time	Medium Time	Long Time
MSE	0.227381705	0.250243248	0.043859639
NMSE	0.925988855	1.042680206	1.361161151
MAE	0.454409816	0.478394991	0.129870423
Min Abs Error	0.208144724	0.199741885	0.055272646
Max Abs Error	0.78607142	0.760175809	0.879882336
Correlation Coefficient (R):	0.316496091	0.223470333	0.675498709
Percentage of output correctly predicted	76.47%	58.33%	100%

Figure 1. Training MSE vs. Epoch.

1 to 7. The extensive use of various network models initially helped to observe that the Multi-layer perceptron was giving consistently better performance accuracy. This agrees with the literature as stated earlier. Statistical performance measures showing accuracy are discussed below.

A model based on the chosen multi layer perceptron having three hidden layers and four processing elements per layer was chosen.

4. Results and discussion

After the training and cross validation, the network was tested with the test data set and the following results were obtained. As shown in the confusion matrix, Table 4 the network was able to attain an accuracy of 77% for short, 58% for medium and 100% for the long classification. This translates to an accuracy of about 70% overall for the Artificial Neural Network which is a fair performance going by similar results from the literature [5,15]. For instance, an ANN achieved 70% accuracy in predicting cocoa production output in Nigeria [1]. Figure 1 indicates that the model training mean square error averages 0.25 after 200 epoch.

5. Conclusion

The study led to the development of an Artificial Neural Network model that can predict the duration of downtime associated with various breakdowns or maintenance activities of a willet-coding machine. A total of six input variables and three output variables were used for the model which achieved an accuracy of over 70% in predicting the expected downtime. Also an initial evaluation of this approach on overall production capacity indicates improvement on machine throughput and the company's ability to meet due date. We recommend that similar studies be carried on other production facilities incorporating input factors such as the ambient conditions.

References

[1] Adefowoju, B. S. and Osofisan, A. O., 2004, *Cocoa Production forecasting Using Artificial Neural Networks.* International Centre for Mathematics and Computer Science Nigeria. ICMCS, 117-136.

[2] Aloba O. O., Oluleye, A. E. and Akinbinu, A. F., 2002, Preventive maintenance planning for a fleet of vehicles using simulation. *Nigerian Journal of Engineering Management*, 3(4), 7-12.

[3] Antognetti, P. and Milutinovic, V., 1991, *Neural Networks: Concepts, Applications, and Implementations.* Volumes I-IV, Prentice Hall, Englewood Cliffs, NJ.

[4] Artana, K. B. and Ishida, K., 2002, Spreadsheet modeling of optimal maintenance schedule for components in wear-out phase. *Reliability Engineering & System Safety*, 77(1), 81-91.

[5] Baesens, B., Van Gestel, T., Stepanova, M., Van den Poel, D. and Vanthienen, J, 2005, Neural network survival analysis for personal loan data. *Journal of Operation Research Society*, 56(9), 1089-1098.

[6] Bobos, A. G. and Protonotarios, E. N, 1978, Optimal system for equipment maintenance and replacement under Markovian deterioration. *European Journal of Operational Research*, 2 , 257-264.

[7] Brouwers, J. H., 1986, Probabilistic descriptions of irregular system downtime. *Reliability Engineering*, 15(4), 263-281.

[8] Emuoyibofarhe, O. J., Reju, A. S. and Bello, B. O., 2003, Application of ANN to optimal control of blending process in a continuous stirred tank reactor (CSTR). *Science Focus; An International Journal of Biological and Physical Science.* 3, 84-91.

[9] Hertz, J. K., 1991, Introduction to the Theory of Neural Computation. Addison Wesley, NY.

[10] Javadpour, R. and Knapp, G. M., 2000, A fuzzy neural network approach to machine condition monitoring. *Computers and Industrial Engineering*, 45(2), 323-330.

[11] Kiureghian, A. D., Ditlevsen, O. D. and Song, J., 2006, Availability, reliability and downtime of systems with repairable components. *Reliability Engineering & System Safety.* available at: doi:10.1016/j.ress.2005.12.003 .

[12] McCluskey, 2002, Types of ANN. available at: http://scom.hud.ac.uk/scomtlm/book/node358.html

[13] Olmsted, D. D., 1999, History and principles of neural networks. available at: http://www.neurocomputing.org/History/history.html .

[14] Principe, J. C., Euliano, N. R. and Curt Lefebvre, W., 2000. *Neural Networks and Adaptive Systems.* John Wiley & Sons, N.Y.

[15] Veiga, J. F., Lubatkin, M., Calori, R., Very, P. and Tung, Y. A., 2000, Using neural network analysis to uncover the trace effects of national culture. *Journal of International Business Studies*, 31, 223-238.

Appendix I: AN overview of artificial neural network

An Artificial Neural Network, (ANN) which imitates the human brain in problem solving, is a more general regression approach capable of capturing all forms of data relationships as well as automatically modifying output in a case of additional information [12,13]. As in conventional regression analysis, the input data x_j are multiplied by weights, but the sum of all these products forms the argument of a hyperbolic tangent. The output y is therefore a nonlinear function of x_j, and combining many of these functions increases the available flexibility. So there is no need to specify a function to which the data are to be fitted. The function is an outcome of the process of creating a network and the network is able to capture almost arbitrary relationships. One desirable feature of the network models is that they are readily updated as more historical data become available. That is the models continue to learn and extend their knowledge base. Thus ANN models are referred to as adaptive systems. The ANN has continued to find applications such as pattern discovery [15], process control [8], financial system management [5], medical diagnosis and other areas [3,10].

Generally a neural network consists of n layers of neurons of which two are input and output layers, respectively. The former is the first and the only layer which receives and transmits external signals while the latter is the last and the one that sends out the results of the computations. The n-2 inner ones are called hidden layers which extract, in relays, relevant features or patterns from received signals. Those features considered important are then directed to the output layer. Sophisticated neural networks may have several hidden layers, feedback loops, and time-delay elements, which are designed to make the network as effective as possible in discriminating relevant features or patterns. The ability of an ANN to handle complex problems depends on the number of the hidden layers although recent studies suggest three hidden layers as being adequate for most complex problems [1].

There are feed-forward, back-propagation and feedback types of network depending on the manner of neuron connections. The first allows only neuron connections between two different layers; the second has not only feed-forward but also 'error feedback' connections from each of the neurons above it. The last shares the same features as the first, but with feedback connections, that permit more training or learning iterations before results can be generated.

ANN learning can be either supervised or unsupervised. In the supervised learning the network is first trained using a set of actual data referred to as the training set. The actual outputs for each input signal are made available to the network during the training. Processing of the input and result comparisons are then done by the network to get errors which are then back propagated, causing the system to adjust the weights which control the network [3,9]. In unsupervised learning, only the inputs are provided, without any outputs: the results of the learning process cannot be determined. This training is considered complete when the neural network reaches a user defined performance level. Such networks internally monitor their performance by looking for regularities or trends in the input signals, and make adaptations according to the function of the network. This information is built into the network topology and learning rules [8,14].

Typically, the weights are frozen for the application even though some network types allow continual training at a much slower rate while in operation. This helps a network to adapt gradually to changing conditions. For this work, the supervised training is used because it gives faster learning than the unsupervised training.

In supervised training, the data is divided into three categories: the training, verification and testing sets. The training set allows the system to observe relationships between input data and outputs. In the process, it develops a relationship between them. A heuristic states that the number of the training set data should be at least a factor of 10 times the number of network weights to adequately classify test data [14]. About 60% of the total sample data was used for network training in this work.

The verification set is used to check the degree of learning of the network in order to determine if the network is converging correctly for adequate generalization ability. Ten samples (10% of the total sample data) were used in this study. The test/validation set is used to evaluate the performance of the neural network. About 30% of the total sample data served as test data in this work.