Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

On the limitations of neural network techniques for analysing cause and effect relationships in manufacturing processes – a case study

R. S. Ransing, M. R. Ransing & R. W. Lewis
Civil and Computational Engineering Research Centre, School of Engineering, University of Wales Swansea, SA2 8PP, UK

Abstract

The cause and effect relationship is complex for many manufacturing processes. The ability to learn causal relationships from diagnostic examples is extremely useful. This learning ability can help not only to quantify the influence of causes on defects for existing components but also to set up new process, material and design parameters to manufacture new components. A neural network approach that can adapt and learn from past examples, was explored for analysing and quantifying cause and effect relationships. Neural networks use the data to extract any pre-existing relationships between the input and output variables. However, in many real situations very few good quality training examples are generally available. A few limitations of multi layer feed forward neural networks in presence of such noisy, limited and sparse training data sets have been discovered, and because of such limitations it has been found difficult to use this technique as a robust tool by end users in a foundry environment or any other manufacturing industry to analyse cause and effect relationships.

1 Introduction

The diagnosis of defective castings has always been a center of attention in the manufacturing and the research community. The "*cause and effect*" relationship in castings is highly complex and non-linear. The diagnostic problem, therefore, was defined as analyzing and quantifying the cause and effect relationships and to study whether neural network's 'learning by examples' paradigm could be

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

used to automatically quantify causal linkages. The work presented in this paper is a part of the ongoing research on the analysis of cause and effect relationships.

1.1 Current approaches for defect analysis

In any Intelligent Diagnostic System, the causal knowledge representation scheme determines the ability of the system to effectively diagnose problem and also to justify its diagnosis. In casting processes, most of the expertise is generally gained, over a period of years, by trial and error. In the late seventies the focus of research for improving the die casting quality was greatly on the experimental side. Statistical techniques such as design of experiments, factorial analyses, and control charts were used to analyse the relationships between causes and defects [1-3]. The influence of design, process and material parameters on the quality of castings was correlated with these analyses. The rule based, deterministic expert systems associated with backward chaining as a reasoning mechanism, have also been developed for casting defect analysis [4, 5]. This technique aroused some interest in the late eighties and early nineties and is also sometimes referred to as knowledge based systems. In these intelligent casting defect analysis applications, the handling of uncertainty [6] in the rules is considered through certainty factors as used in MYCIN [7] - one of the earlier and commercially successful expert systems. Many other expert systems were developed for casting analysis [4, 6, 7, 8, 9]. These expert systems are essentially based on a rule based architecture [10]. Generally, the causal relationship is highly complex and inter-linked i.e. a cause influences a number of defects and each defect is influenced by a large number of causes. A comprehensive rule base would then require a large number of rules leading to other computational problems such as efficiency of the computation, consistency in the rules etc. Also, the fuzzy sets or rule based expert system approach requires the knowledge of the degree of influence of the related cause on the occurrence of each defect as an input to the system. To our experience, generating such a probability distribution for the entire relationship is extremely difficult if not possible.

1.2 Neural networks for defect analysis

Recently, the neural network technology has gained more popularity because, unlike the rule-based approach this technology offers a convenient computational tool that can adapt and learn from past examples. This technique can also be used to quantify complex and highly inter-linked causal relationships. Smith and her co-investigators [11,12] have done considerable research in this direction. Smith et. al. have shown that the back-propagation neural network can be successfully applied for quality control applications. Neural networks were also used for the diagnosis of hydraulic forging presses [13]. Martinez et. al. [14] have investigated its application to relate process conditions to the probable quality rating of casting. This predictive analysis was done for a slip casting process. Spelt et. al. [15] have used neural networks for classifying power plant sensor

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

data and coupled this with an expert system for diagnostic purposes. Zhang and Huang [16] have presented a state-of-the-art survey of neural network applications in manufacturing. These include applications of neural networks for engineering design, process planning, in solving scheduling problems, process modelling and control, in monitoring and diagnosis and quality assurance. The semantically constrained neural network approach proposed by Ransing and Lewis [17] has a modified three layered 'feed forward' network architecture. It constrains the connectivity of nodes according to the 'defect-metacause-rootcause' relationship [23].

In this paper, the "Learning from examples" ability of neural network techniques has been studied in detail and the issues involved in quantifying cause and effect relationships for manufacturing processes have been discussed and the areas of improvements required in the current understanding of the neural network modelling have been highlighted.

2 Neural network modelling

The standard multi-layer feed forward neural networks have been considered that have an input layer of neurons, a hidden layer of neurons and an output layer of neurons. The diagnostic knowledge, which the network learns is stored in terms of weights, i.e. numeric values associated with the links connecting network nodes. An adjustment of weights during the back-propagation training algorithm allows the network to learn and recognise patterns.

When using the back-propagation algorithm to train a multilayer neural network, the designer is required to arbitrarily select parameters such as the network topology, the initial weights, the initial biases, the learning rate, the activation function, the gain value used in the activation function and the momentum constant. An activation function is used for limiting the amplitude of the output of a neuron. For this study, a smooth and continuous logistic sigmoid activation that outputs values within the range 0 to 1 is chosen. A unit value of gain has been chosen for training the network. Weights and biases have been initialised to small random numbers. Most of the neural network applications used in the manufacturing field have used a constant value for learning rate between 0.3 and 0.6. The learning rate determines the magnitude of weight change along the chosen direction in weight space. If a larger value is chosen for the learning rate, it may speed up the learning process, however, it may also result in large changes in the synaptic weights that make the network unstable. Smaller values of learning rates may avoid oscillatory changes in weights, however, at the cost of excessive computational time. A general rule is to use the largest learning rate that works and does not cause oscillation. In this study, a learning rate of 0.3 and momentum of 0.4 has been chosen. In addition to the learning rate parameter, the momentum term also influences the convergence speed of the training procedure. The momentum constant, μ, is a fraction which determines the proportion of weight change in the previous iteration that is added in the current weight change. This ensures a smooth change in weights by minimizing oscillations leading to a better convergence rate. Several attempts

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

have been made to explicate the influence of momentum term upon the back propagation [18] to suggest an optimal value for the term [19, 20]. As discussed before, the larger the learning rate η, the larger is the change in the weights. However, by using a moderate value of the learning rate and adding a momentum term filters out the oscillations [21, 22]. It has also been observed that learning rate and momentum constant have a significant impact on the training speed, but not on the generalization ability [19].

An improper choice of any of the above parameters can result in very slow convergence or even network paralysis where the training process comes to a virtual standstill.

The processing at each neuron is almost the same and the final network architecture is generally determined by a trial and error method. The number of hidden layers and also hidden nodes in each layer determine the non-linearity of the mapping function in the input and output space. The generalisation ability of the network is directly determined by the number of hidden nodes in the network.

2.1 Input and output of network models

As the cause and effect relationship is highly inter-linked, the occurrence and non-occurrence of defects determine the occurrence and non-occurrence of causes. The activations for all the nodes are scaled from 0 to 1. A value of zero corresponds to the non-occurrence of the nodes and a value of one corresponds to the occurrence of the node and the fractions denote the strength of occurrence of the corresponding node.

For a manufacturing process, the rejection data consists of the number of defective components for each defect type. A component showing two defect types is counted under both defect types. The defect type with the maximum number of components is assigned a unit value and the rest of the defect types were proportioned accordingly to the scale from zero to one. In this way, the relative strength of each defect is calculated. This forms an input vector to the network. The activations in the output nodes, each corresponding to a particular cause, generate the diagnostic output for the network model.

2.2 Performance of traditional multi-layer neural network on a real data set

The major difficulty that the authors experienced in using the neural network technology was in the collection of suitable training examples. Under the practical situation, it is extremely difficult, if not impossible, to get even one hundred representative examples. Some of the examples collected were so similar to each other that their inclusion would have just duplicated the training data set. Most of the examples were also sparsely located in the input space i.e. the spread of examples did not cover the entire input space. In other words, the extrapolation abilities of the network became more important than the interpolation capabilities.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

These difficulties in generating the training set got amplified with the existence of contradicting examples (it is indeed true that on many occasions, two experts in one field do not agree with each other!). The training data set also had noisy data.

The performance of the neural network technique was tested under these practical limitations of generating a good quality training data set and some of the potential weaknesses of this method have been highlighted.

The data was collected from a pressure die casting foundry in the UK. A total of 14 defects were identified and associated with 43 process, material or design parameters. The data was collected for similar components over a period of one year and a total of 60 representative examples were finalised. A belief value in the occurrence of defects was calculated as described previously in Section 2.1. The corresponding belief values representing the occurrence and non-occurrence of associated process, design and material parameters were given by the experts in the foundry. To show the graphical variation of belief surfaces learnt by the multi layer feed forward neural network, three defects identified as 'Porosity', 'Mismakes' and 'Dimensional' were chosen and are represented as defect A, defect B and defect C in Table 1. Sixteen associated process, material and design parameters were identified to create a neural network with three input nodes corresponding to defects 'defect A', 'defect B' and 'defect C' and sixteen output nodes corresponding to the sixteen process, material and design parameters. The neural network was chosen with five hidden nodes. The code for a multi layer neural network has been written by the authors in both Matlab and C++ programming language. The performance of the network was tested and compared using faster optimisation algorithms such as the Quasi-Newton method. In all cases, the network achieved the target error of 0.01 and also performed well on the testing data set.

In all cases, it was discovered that the interpolation abilities of the network were excellent. However, the network consistently demonstrated very poor extrapolation capabilities. The extrapolation abilities of the network can be assessed by visualising a surface describing the variation in belief values in a cause corresponding to two associated defects in each instance. Figure 1(a) plots the surface learnt by the neural network model describing the variation in belief values for the cause 'The position of gate' corresponding to the defects 'porosity' and 'mismake'. It is assumed that the third associated defect 'dimensional' occurs at a very low proportion. The corresponding surface visualised by an expert is shown in Figure 1 (b). The difference can be clearly observed. Figures 1 (c) and (d) plot the surface when the third defect 'dimensional' occurs correspondingly at medium and high proportions. The pattern learnt by the neural network is clearly unrealistic. Similar trend was observed for all other causes and defects.

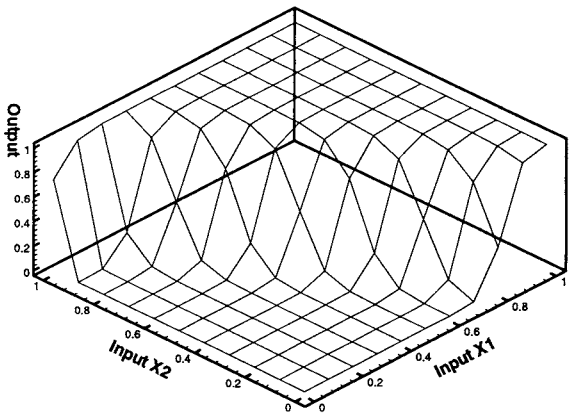


Figure 1a: Output Surfaces generated by the traditional neural network for a cause "The position of gate". For an input value of 0.1 for the "Dimensional" defect. X1: Belief value in the occurrence of "Porosity" defect. X2: Belief value in the occurrence of defect "Mismake" defect. Output: Belief value in the occurrence of "The position of gate".

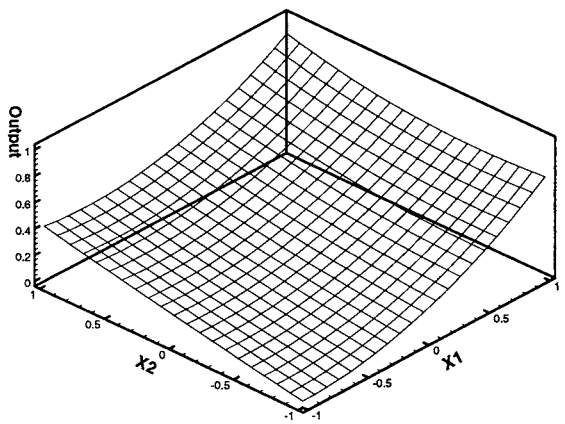


Figure 1b: Output Surfaces generated as visualised by an expert for a cause "The position of gate". For an input value of 0.1 for the "Dimensional" defect. X1: Belief value in the occurrence of "Porosity" defect. X2: Belief value in the occurrence of defect "Mismake" defect. Output: Belief value in the occurrence of "The position of gate".

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

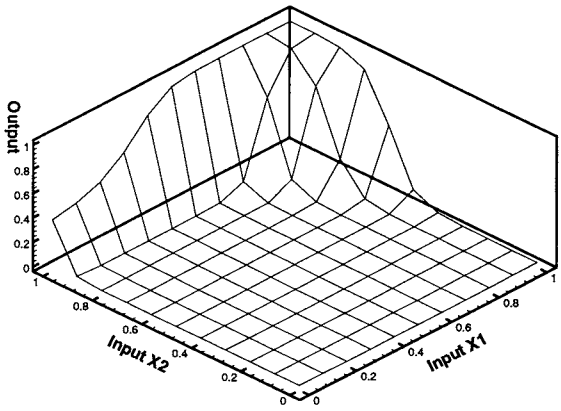


Figure 1c: Output Surfaces generated by the traditional neural network for a cause "The position of gate". For an input value of 0.5 for the "Dimensional" defect. X1: Belief value in the occurrence of "Porosity" defect. X2: Belief value in the occurrence of defect "Mismake" defect. Output: Belief value in the occurrence of "The position of gate".

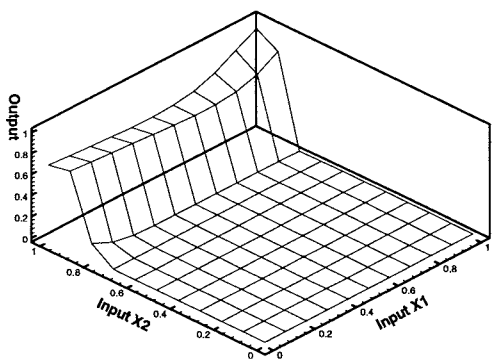


Figure 1d: Output Surfaces generated by the traditional neural network for a cause "The position of gate". For an input value of 0.5 for the "Dimensional" defect. X1: Belief value in the occurrence of "Porosity" defect. X2: Belief value in the occurrence of defect "Mismake" defect. Output: Belief value in the occurrence of "The position of gate".

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

3 Conclusions

In this paper a neural network approach that can adapt and learn from past examples, was explored for analysing and quantifying cause and effect relationships. Neural networks use the data to extract the relationship between the input and output. Data collected for a pressure die casting process has been used to train a multi layer neural network. As seen in many real situations, this data set is also noisy, limited and data points are sparsely located in the input space. After training the neural network and plotting the output a few limitations of neural network for precisely quantifying cause and effect relationships for manufacturing process have been discovered. Extrapolation abilities of neural networks trained on such data sets are found to be very poor as output depends on the location and number of data points as well as the target output of each data point in the training data set. Also, the parameters in the network as well as the network topology are also decided by trial and error method. An improper choice of any of these results in problems such as very slow convergence, network over-fitting, etc.

Our advice for manufacturing industries is that feed-forward neural networks are nothing more than efficient, multi-dimensional, multi-variable, non-linear interpolation algorithms based on least-square fit principles and should not be used like a black box. Collecting good quality training data is the major issue for using this technique in manufacturing diagnosis and it should be used with extreme caution. For many foundries, it is extremely difficult and expensive, if not impossible, to collect the good quality training data as required by neural networks. Research is currently underway to develop a new algorithm in order to overcome this limitation of the neural network modelling.

Acknowledgements

The authors gratefully acknowledge the financial support of EPSRC (GR/R89912) and Rolls Royce Plc to continue this work.

References

[1] Askeland D. and Patel D., Factorial Analysis of Some Variables and Their Optimisation for the Production of Sound Die Casing, *Proc. 9th International Die-Casting Exposition and Congress,* 1977, paper no. G-T77-013.

[2] Wightman D. E., Temperature Control in the Die-Casting Process Through the Use of a Heat Transfer System, *Proc. 9th International Die-Casting Exposition and Congress*, 1977, paper no. G-T77-066.

[3] Takach B. V., Some Aspects of Feed Design for Pressure Die-Casting Dies, *Proc. 10th International Die-Casting Exposition and Congress,* 1979, paper no. G-T79-094.

[4] Creese R. C., Introduction to Expert Systems to Foundry Application, *AFS Trans,* 1988, 96, 443-446.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

[5] Roshen H.Md., Expert System for Analysis of Casting Defects: Cause Module, *AFS trans*, 1989, 97, 601-606.

[6] Er. A., Kondic V., Knowledge-based systems and their application in casting defects control, *International Journal of Cast Metals Research,* 1996, 9(3), 163-182.

[7] Shortliffe E. H., *Computer based medical consultations: MYCIN,* New York: American Elsevier, 1976.

[8] Piwonka T. S., Current and Potential Use of Process modeling for Foundry Process Control, *AFS trans,* 1989, 97, 465-472.

[9] Phelps T. A., Analysis of Internal Unsoundness Casting Defects using Artificial Intelligence Techniques, *AFS Trans,* 1989, 97, 507-512.

[10] Davis R., Buchanan B. G. & Shortliffe E. H., Production rules as representation for a knowledge-based consultation program, *Artificial Intelligence,* 1977, 8, 15-45.

[11] Smith A. E., Predicting product quality with backpropagation: a thermoplastic injection moulding case study, *International Journal of Advanced Manufacturing Technology,* 1993, 8, 252-257.

[12] Smith A. E. & Dagli C. H., Controlling industrial processes through supervised, feedforward neural networks, *Computers and Industrial Engineering,* 1991, 21, 247-251.

[13] Lin H., Yih Y. & Salvendy, Neural network based fault diagnosis of hydraulic forging presses in China, *International Journal of Production Research,* 1995, 33(7), 1939-1951.

[14] Martinez E. E., Smith A. E. & Idanda B., Reducing waste in casting with a predictive neural model, *Journal of Intelligent Manufacturing*, 1994, 5(4), 277-286.

[15] Spelt P. F., Knee H. E. & Glover C. W., Hybrid artificial intelligence architecture for diagnosis and decision-making in manufacturing, *Journal of Intelligent Manufacturing,* 1991, 2, 261-268.

[16] Zhang H. C. & Huang S. H., Aplications of neural networks in manufacturing: a state-of-the-art survey, *International Journal of Production Research,* 1995, 33(3), 705-728.

[17] Ransing R. S. & Lewis R. W., A Semantically Constrained Neural Network for Manufacturing Diagnosis, *International Journal of Production Research,* 1997, 35 (9), 2639-2660

[18] Perantonis S.J. & Karras D.A., An Efficient Constrained Learning Algorithm With Momentum Acceleration, *Neural Networks*, 1995, 8(2), 237-249.

[19] Sato A., An Analytical Study of the Momentum Term in a Back Propagation Algorithm, *Proceeding of ICANN91*, 1991, 617-622.

[20] Holger R.M. & Graeme C.D., The Effect of Internal Parameters and Geometry on the Performance of Back-Propagation Neural Networks, *Environmental Modelling and Software,* 1998, 13(1), 193-209.

[21] Dai H. & MacBeth C., Effects of Learning Parameters on Learning Procedure and Performance of a BPNN, *Neural Networks,* 1997, 10(8), 1505-1521.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

[22] Qian N., On the Momentum Term in Gradient Descent Learning Algorithms, *Neural Networks,* 1999, 12, 145-151.

[23] "ICADA: Intelligent Computer Aided Defect Analysis for Castings", RS Ransing, MN Srinivasan and RW Lewis, *Journal of Intelligent Manufacturing*, vol. 6, no. 1, 29-40, 1995.