Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 5
Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima, Soo-Young Lee, and Xin Yao (Editors)

THREE NEURAL NETWORK CASE STUDIES IN BIOLOGY AND NATURAL RESOURCE MANAGEMENT

Sandhya Samarasinghe[1]
Don Kulasiri[1]
Channa Rajanayake[2]
Meegalle Chandraratne[3]

[1] Centre for Advanced Computational Solutions, [2] Applied Computing, Mathematics and Statistics Group, [3] Molecular Biotechnology Group, Lincoln University, Canterbury, New Zealand

ABSTRACT

This paper presents 3 NN case studies. In the first, fracture toughness of wood was predicted using an expanded MLP network from experimentally measured crack angle, stiffness, density and moisture content. The data is characterized by noise but the model produced physically meaningful nonlinear trends with an R2 value of 0.67. In the second study, hydraulic conductivity (K m/day) was estimated from ground water solute concentration data collected for a range of K values. Four separate NN needed to be developed for four sub-ranges of K to reduce error. In order to determine the appropriate range of K for a particular system, concentration data were clustered into 4 groups using SOM. The hybrid-model was applied to an experimental aquifer and only 10% difference was found between experimental and NN estimations of K. In the third study, digital images of lamb chops were used to collect values for 102 geometric and textural variables for meat grading. Principal Component analysis reduced the variables to twelve. Three- and 4- layer MLP networks and Discriminant Function Analysis (DFA) were performed on the data and the classification accuracy from 3-layer MLP was 83% and was 12% better than that from DFA.

CASE STUDY 1: PREDICTING FRACTURE STRENGTH OF WOOD

Wood is a construction material that is also a complex biological material. Its mechanical properties are highly variable due to variable cellular structure that interacts with moisture [1]. In this study, fracture toughness of wood as influenced by material and physical properties was studied using neural networks. Fracture toughness indicates the resistance to failure of a material containing crack-like defects. The properties used to predict fracture toughness were elastic stiffness, dry density, moisture content and angle of orientation of the crack with respect to the failure path.

1.1. Objectives

The aim of this study was to assess the capability of neural networks to capture general trends to predict fracture toughness of wood from variable and noisy input data.

1.2. Data

Data was obtained from experiments conducted on wood specimens containing a crack. For each specimen, fracture toughness, elastic stiffness, dry density, moisture content and crack angle to the natural failure path were determined. There were a total of 35 records of data.

Fracture strength was plotted against each of the four variables and the four variables were plotted against each other and the plots and the correlation matrix was analysed. Fracture strength was most positively correlated with crack angle, and to a lesser extent with Young's modulus and density, and negatively correlated with moisture content. Plots showed a large scatter as is usually the case.

1.3. Neural Network Modelling

Three approaches were taken to modelling: (1) Multiple linear regression (MLR), (2) Standard multi-layer perceptron network (MLP) and (3) Expanded multi-layer perceptron (WARD^{TM}).

The MLR was conducted on Minitab (Release 13.1) and resulted in an R^2 value of 0.52. This assumes a linear relationship between the output and inputs. To further improve the results and search for possible nonlinear relationships, the neural network approach was attempted [2]. The dataset was randomly divided into training (80%), testing (10%) and validation (10%) datasets. Testing dataset was used to prevent overfitting and validation set was used validate the developed model.

Neural network models were developed on NeuroShell2 (Ward Systems Group, USA, 1997). For standard MLP, many parameter combinations were tested. However, a significant improvement of results was not achieved. Therefore, the WARD network was tested on the data. This network has a single hidden neuron layer consisting of two slabs. Each slab contains a set of neurons represented by a particular transfer function. Thus, a two hidden slabs represent two transfer functions.

There was an immediate improvement on the predictions from the WARD network and therefore, various possibilities for transfer functions, number of hidden neurons per slab, data scaling functions, initial random weights, and two learning laws: steepest descent and TurboProp were tested. Furthermore, since the dataset was small (35 records), modelling

was done on 5 different sets of training, testing and validation data randomly extracted from the original dataset. The five models provided very consistent results indicating that they have captured a general trend in the data. The resulted average R^2 was 0.67. The final model had the following features:
Transfer functions: slab 1—Gaussian Complement, Slab 2—Gaussian Complement; Number of Neurons: 8 per slab; Learning law: Steepest Descent; Learning Rate=0.8, Momentum =0.1;
Input scaling: <-1,1>.

The Gaussian and Gaussian Complement transfer functions provided much better results than logistic and tanh etc. functions. We attribute this to the Gaussian function being sensitive to the average and Gaussian Complement to the extremes thereby more efficiently covering the total range of input data that had distributional characteristics.

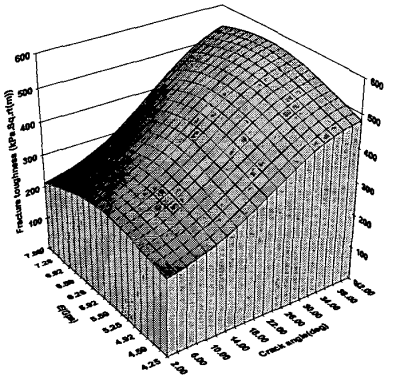

Fig. 1(a)- Predicted trend for the fracture toughness, stiffness (E) and Crack angle relationship

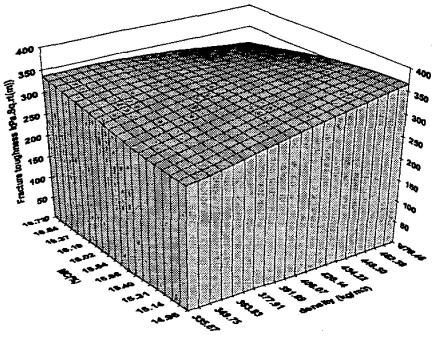

Fig. 1(b)-Predicted trend for the fracture toughness, density and moisture content relationship

A sensitivity analysis was done using the developed model to assess the contribution of each input variable by varying it from the minimum and maximum value while holding others at their mean. It showed that the crack angle contributed the most (40%) followed by stiffness (20%), dry density (19.5%) and moisture content (18.9%).

The developed model was also used to study the nature of the influence of the four variables on fracture toughness. Figure 1 shows the predicted relationship between Fracture toughness, crack angle and Stiffness (E). It shows that the fracture toughness increases significantly and nonlinearly with the crack angle and to a smaller extent with the stiffness. Figure 1(b) shows the influence of moisture content and dry density which indicates that the fracture toughness increases with density but the moisture influence is somewhat inconsistent. For example, the expected trend of fracture toughness decreasing with increasing moisture was observed for the higher densities but not at the lower density range. We attribute this to insufficient data.

1.4 Conclusions

An expanded MLP network consisting of two sets of neurons represented by Gaussian and Gaussian Complement transfer functions in the hidden layer was developed to predict fracture toughness of wood from crack angle, stiffness, moisture content and density. The model accuracy (R^2=0.67) was reasonable considering the highly variable nature of the input data. This model was superior to MLR and standard MLP networks. The model predicted acceptable trends.

CASE STUDY 2: INVERSE PROBLEM IN HYDROLOGY

Groundwater contamination is a widespread issue and has been the subject of numerous studies. In these studies, we simulate or represent the interested system by a mathematical model (by excitation and response relationship) for forecasting and management problems. In the process of developing the models, we introduce the parameters, which we consider to be attributes or properties of the system [3]. Presently accepted direct methods for parameter estimation such as pumping tests are subjected to numerous randomness issues and fail to produce reliable results [4]. Furthermore, such field scale experiments can be expensive and cannot be expected to give a representative parameter value for large heterogeneous aquifers. Conversely, often we are interested in modelling quantities such as the depth of water table and solute concentration. This is because they are directly relevant to environmental decision making, and we measure these variables regularly and relatively more cheaply. Furthermore, we can continuously monitor these decision (output) variables in many situations. If the dynamics of the system can reliably be modelled, we can expect the parameters estimated based on the observations may give us more reliable representative values than those obtained from traditional field or laboratory tests.

2.1 Objectives

The aims of the study were to: (a) develop an inverse methodology to estimate a groundwater parameter, namely

hydraulic conductivity (K, m/day), by using distributed contaminant concentration values of saturated groundwater flow by means of ANN, and (b) test the model on an experimental aquifer.

2.2 Neural Network Modelling

We used two-dimensional deterministic advection–dispersion equation to model the 10 m x 5 m 2-D groundwater system to obtain data [5]. Initial concentration value of 100 ppm was considered as a point source at middle of the header boundary of the aquifer and the same source was maintained at the boundary throughout the 10 day time period considered. Exponentially distributed point source concentration values along the longitudinal and lateral directions were considered as the initial conditions of other spatial coordinates. 800 datasets were generated for different K values that range from 40 to 240 m/day. Each dataset comprised 50 inputs of solute concentration values for 10 distinct spatial locations for 5 different time intervals, and 1 corresponding output variable of K.

A 3 layer MLP network was utilised to train the network to find the relationship between output, K, and the associated concentration values. The dataset was divided into two categories, 80% of them were used for training and the rest was utilised for testing. We applied scale functions of none, logistic and logistic for input, hidden and output layers, respectively. The network parameters used were; learning rate = 0.1, momentum = 0.1, initial weight = 0.3. After a number of trial and error tests, it was found that the optimum results could be achieved by 20 hidden neurons. The network reached the stopping criterion of average error on test set, fixed at 0.000002, in less than 2 min in a 1GHz personal computer with performance measurements of the coefficient of multiple determination, R^2 = 0.9999 and the square of the correlation coefficient, r^2 = 0.9999. The network that produces best results on the test set is the one most capable of generalising and this was saved as the best network.

Having completed the successful training, another dataset was employed to test the prediction of the estimated parameter. We made use of the same model to generate 800 new data values, however, initial concentration was randomly changed by up to ±5% and another up to ±5% noise was arbitrarily added to all concentration input values. The reason for adding the noise is to simulate the real world problem of erratic behaviour of aquifers. The validation data revealed that the network error increased with the magnitude of K with mean square error (MSE) of 5.63% and maximum error of 22.45 m/day (Table 1), which may not be acceptable in most practical cases. Since the objective range of K was fairly large (40 –240 m/day), the accuracy of the approximation tends to decrease for larger K values. Therefore, we developed 4 NNs for four smaller sub-ranges of K: (i) 40-90, (ii) 90-140, (iii) 140-190 and (iv) 190-240 m/day. Table 1 shows that accuracy of the estimates improved considerably with use of smaller target ranges. As an example, the maximum error for the 190-240 range, where the error was over 22 m/day with a larger target range, has been reduced by about 90% to 2.98.

We then investigated the robustness of the model in terms of its ability to predict estimates in different random conditions of the system. One of them may be initial value problem. Hence, we examined the stability of the model for different initial values for the range of K between 190 – 240 m/day. The point source value of the initial concentration as well as all other resulting initial values of the system were changed from –50% to 50% and the parameter was estimated.

Table 1. Statistics of estimated error for different ranges of K with up to ±5% difference in initial value and up to ±5% noise in observations.

Error	K Range (m/day)				
	40-240	40–90	90–140	140–190	190-240
Max	22.45	1.88	2.23	2.99	2.98
Mean	8.03	0.27	0.38	0.36	0.39
StDev	5.15	0.32	0.41	0.49	0.47
MSE	5.63	0.11	0.12	0.18	0.18

Furthermore, to illustrate the heterogeneity of the aquifers, up to ±5% extra noise was randomly added to the concentration values. Results showed that the estimation error increased with the magnitude of the initial value variation. However, maximum error for 50% variation of initial value was only around 10 m/day, which was about 1% of percentage absolute error, and the results are dependable even at higher noise levels.

Random boundary conditions and irregular porous structure can result in erratic distribution of flow paths. Therefore, solute concentration spreads could be highly stochastic. We addressed this issue by extending the investigation of robustness by adding different level of randomness to the concentration values. Once more the estimates showed that the ANN model was reasonably stable and percentage absolute error for fairly high randomly added noise of 50% was only 0.93 % with maximum error of 6 m/day.

Above results and investigation of validity showed that satisfactory results could be obtained by using reasonably smaller target range of parameters. Therefore, in order to apply the models to a real aquifer it was necessary to find the appropriate sub-range of K that represents the actual aquifer. However, in real world problems the prior knowledge of the system is limited. In order to facilitate this, a Self Organising Map (SOM) was used to determine the sub-range of K [6].

SOM has the ability to cluster the data of similar attributes into lower dimensions. We employed SOM to cluster 800 x 50 dimension noisy dataset with parameter range of 40 – 240 m/day into four different categories. The Supervised SOM network architecture successfully categorised four different groups with 201, 200, 197 and 202 data patterns in each cluster as shown in Figure 2. SOM put data into categories with high accuracy with few exceptions, which can be expected with noisy data, at the boundaries of the parameter ranges. Then we created and fed 10 different test datasets with the same number of input variables (50) into the trained SOM and it accurately identified the correct parameter range for all the datasets.

2.3 Application of hybrid-model to an experimental aquifer

We applied the hybrid inverse approach (SOM and MLP) to estimate hydraulic conductivity of an experimental aquifer. We obtained the data for this investigation from a large, confined, artificial aquifer which is being used for contaminant transport tests at Lincoln University, New Zealand. This aquifer is 9.49 m long, 4.66 m wide and 2.6 m deep, and porous media is sand. Although, initial conditions, other parameters and the subsidiary conditions are somewhat known, we had to conduct a fairly tiresome, "trial and error" exercise to replicate the aquifer. 800 data patterns were generated for the hydraulic conductivity range of 80 to 280 m/day. Each pattern consisted of 100 concentration input variables for 10 spatial locations for 10 different time intervals. Then we used SOM (80% data for training and 20% for testing) to classify the input values into 4 clusters. Then we fed the actual aquifer data into the trained network to select the appropriate sub-range of K. Here, the trained network determines the cluster that most resembles the input concentration vector. It was determined that the aquifer parameter should be within the second cluster (130 – 180 m/day). Based on this information we generated a separate dataset for the specified range and trained an MLP network with associated *K* values.

The estimate given by the trained ANN was 152.86 m/day. The experimental value of hydraulic conductivity, *K*, was found to be 137 m/day, which was calculated by calibration tests conducted by aquifer testing staff. In these experiments, they have assumed that the aquifer is homogeneous. The difference between two estimates is only 10.37%. Considering the assumptions of homogeneity made by the aquifer researchers and other possible human errors, they stated that the estimate from ANN model was reasonable and acceptable.

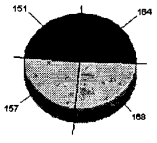

Fig. 2. Distribution of clusters by SOM.

CASE STUDY 3: COMPUTER VISION FOR MEAT GRADING: STATISTICAL AND NEURAL NETWORKS APPROACHES

Computer vision is being used with much success in the automation of many food industry applications, particularly in quality control. It has enormous potential for evaluating meat quality as image-processing techniques can quantitatively and consistently characterize complex geometric, colour and textural properties.

Visual assessment has become the principal component of several meat classification and grading systems. Furthermore, the meat industry, in response to consumer demand for products of consistent quality, is placing more and more emphasis on quality assurance issues. Instrument grading of animal carcasses has been studied to meet the demand for increased accuracy and uniformity of meat grading.

The assignment of lamb carcasses to specific quality grades has been an integral part of the New Zealand lamb classification system. The current classification is based solely on measures of carcass weight and fatness. There are five weight (A, L, M, X or H) ranges and five fatness (A, Y, P, T or F) classes [7]. The final assigned grade is a combination of fatness class followed by weight class (e.g. YM, PX etc).

The traditional statistical modelling methods have limited capability to model an unknown non-linear relationship between the features of the image and the attributes of the meat. Neural network (NN) could be a promising alternative method in prediction modelling when there exists an unknown non-linear relationship between the inputs and outputs of the prediction model.

3.1. Objectives

The objectives of the study were: a) to develop a NN model to evaluate lamb carcass grade using data extracted from lamb chop images; b) to compare the NN with Discriminant Function Analysis (DFA) approach.

3. 2. Methods

3.2.1 Samples

The data was collected from 160 digital images of lamb chops. Six common lamb grades (YM, YX, PM, PX, TH, FH) were selected for this analysis.

3.2.2. Imaging system

The imaging system consisted of a Digital Camera (Sony colour digital camcorder DSR-PD150P), Lighting system (RSX Copy Stand with two sets of RB 5004 HF Copy Lighting units, Kaiser, Germany), Personal Computer (850 MHz AMD Athlon processor, with 512 MB RAM), Image processing and analysis software (Image-Pro Plus, Media Cybernetics, USA).

3.2.3. Image Capture

Surface moisture was removed with a paper towel prior to image capture. For imaging, lamb chops were placed flat on a non-glare black surface and illuminated with standard lighting. The still images of lamb chops were later transferred to the PC for storage and analysis. The images included lean area, marbling, subcutaneous fat, intermuscular fat and bone.

3.2.4. Image Processing and Analysis

Image processing and analysis were accomplished in the Windows 98 environment using Image-Pro Plus. The images were first segmented into lean (dark) and fat (light) areas. Thresholding was done through trial and error by observing and selecting the best value. Initial values for thresholding were selected from the plot of pixel intensities. A total of 12 image geometric (thickness and area) variables (*lean area, marbling area, subcutaneous fat area, lean ratio* (lean area/(lean area + marbling area)), *number of marbling specks, 3 measurements of subcutaneous fat thickness (average, maximum, minimum), fat thickness at 11 cm from the midline of the carcass and 3 measurements of fat thickness between 10 and12 cm from the midline of the carcass (average, maximum, minimum*) were measured.

3.2.5. Textural Properties

Texture is the term used to characterize the surface of a given object or phenomenon. It is one of the main features used in the image processing. Grey level co-occurrence matrix (GLCM) defined by Haralick, et al, [8] was used to extract texture features. GLCM is based on the estimation of second order joint conditional probability density functions. The number of texture parameters calculated from grey level co-occurrence matrix was 90 [8,9].

3.3. Data Analysis

Image data was analysed using NN and DFA procedure was used for comparison. Total number of inputs to the NN was 102. With 6 outputs this probably could lead to over-parameterisation and is a serious problem in meat quality prediction. Therefore, an effective method for reduction of feature dimension of the prediction model was desired. Principal Component Analysis (PCA) and Cluster Analysis together with correlation coefficients and analysis of variance (ANOVA) were used to reduce the dimensionality of data. Statistical analysis was performed with Minitab (release 13.1, Minitab Inc.). According to the results of PCA, 96% of the total variance of image geometric variables can be condensed into six variables. In a similar way, 90 texture variables were condensed to six, whilst retaining 99.5% of the total variance.

The final set of six geometric variables were: *lean area, marbling area, fat area, lean ratio, subcutaneous fat thickness average and 11 cm fat thickness,* and six texture variables were: *homogeneity, entropy, contrast, cluster prominence, difference entropy and information measures of correlation 2.*

3.4 Neural networks and Statistical Analysis

Neural network analysis was performed with NeuroShell® 2 (Ward Systems Group, USA). The multi-layer perceptron (MLP) with back propagation algorithm was performed to train the network. The 12 variables in the reduced set were used as inputs and the grades were used as outputs. To avoid any hierarchy among grades, the symbols representing six grades were translated into six variables, one for each grade, that is coded 1 for belonging to the grade and 0 for not belonging. Several different neural networks and learning schedules were tested to select the best neural network.

Test (20%) and validation (20%) sets were randomly extracted from the original data. Linear activation functions were used for input nodes and logistic functions were used for hidden and output nodes. Neural networks with number of hidden layers (1 to 3) and hidden nodes (12 to 22) were trained. Different combinations of learning rates (0.1 to 0.5), momentums (0.1 to 0.5) and initial weights (0.2 to 0.6) were tested. Different weight updates (vanilla, momentum and turboprop) and different pattern selections (rotation or random) were also tested.

In processing data file through the trained neural network the highest output was set to 1 and the others to 0. Single hidden layer NN with 18 hidden nodes, turboprop weight update and rotation pattern selection produced highest classification (83%). The learning rate, momentum and initial weight used were 0.1, 0.1 and 0.5 respectively. The minimum average error was 0.021 and this average error was achieved after 10000 epochs. The highest classification from 2 hidden layers network was only 70%. The results from neural network models are shown in Table 2 which shows a remarkable accuracy that varies somewhat with the grade (Table 2).

DFA finds a set of linear combinations of the variables, whose values are as close as possible within groups and as far apart as possible between groups. The linear combinations are called discriminant functions. DFA was performed using the reduced set of six geometric and six texture and the analysis was carried out using linear discriminant function and without cross validation. The accuracy of prediction was 71

Table 2. Results of classification of meat grades

	Analysis method	No of images correctly classified (%)						
		YM	YX	PM	PX	TH	FH	total
1	DFA	77.4	68.0	73.3	63.0	80.0	100	71.0
2	4 layer MLP	80.6	78.7	13.3	70.4	0.0	100	70.0
3	3 layer MLP	77.4	86.7	86.7	81.5	60.0	100	83.0

3.5. Conclusions

The results indicate that the data extracted from images of lamb chops can be effectively used to predict the lamb carcass grades. The accuracy of prediction using the NN approach and DFA was 83% and 71% respectively. The NN approach was superior to the DFA method and improved the overall classification accuracy by 12%.

4. REFERENCES

[1] Bodig, J. and Jayne, B.A. *Mechanics of wood and wood composites,* Van Nostrand Reinhold Company, Inc. 1982.

[2] Smith, M. *Neural networks for Statistical Modelling,* 1996.

[3] Bear. J., *Hydraulics of groundwater.* McGraw-Hill Inc, Israel, 1979.

[4] Unny. T.E., Stochastic Partial Differential Equations in Ground Water Hydrology - Part 1: *Journal Hydrology and Hydraulics*; 3:135-153, 1989.

[5] Fetter. C. W. - *Contaminant Hydrogeology,* Prentice-Hall Inc., New Jersey, 1999.

[6] Kohonen, T., Self-organized formation of topologically correct feature maps, *Biological Cybernetics, 43,* 59-69, 1982.

[7] New Zealand meat guide to lamb and mutton carcass classification, New Zealand Meat Board, Wellington, New Zealand, 1992 (Reissued 1995).

[8] Haralick, R.M., Shanmugan, K. and Dinstein, I. "Texture features for image classification", *IEEE Trans. on Systems, Man and Cybernetics,* SMC-3 (1), 610 - 621, 1973.

[9] Unser, M. "Sum and difference histograms for texture classification", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* PAMI-8 (1), 118 - 125, 1986.