

Neural Networks in Seismic Discrimination

Farid U. Dowla

This was prepared for submittal to the
*North Atlantic Treaty Organization
Advanced Study Institute Meeting
Alvor, Algarve, Portugal
January 23-February 2, 1995*

January 1995

The logo of Lawrence Livermore National Laboratory, featuring a stylized 'L' and the text 'Lawrence Livermore National Laboratory' arranged in a chevron shape.

Lawrence
Livermore
National
Laboratory

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

NEURAL NETWORKS IN SEISMIC DISCRIMINATION

Farid U. Dowla
Lawrence Livermore National Laboratory
Livermore, California 94550

1. Abstract

Neural networks are powerful and elegant computational tools that can be used in the analysis of geophysical signals. At Lawrence Livermore National Laboratory, we have developed neural networks to solve problems in seismic discrimination, event classification, and seismic and hydrodynamic yield estimation. Other researchers have used neural networks for seismic phase identification. We are currently developing neural networks to estimate depths of seismic events using regional seismograms. In this paper different types of network architecture and representation techniques are discussed. We address the important problem of designing neural networks with good generalization capabilities. Examples of neural networks for treaty verification applications are also described.

2. Introduction

The computational paradigm of neural networks has several attractive features for solving problems in monitoring a Comprehensive Test Ban Treaty. A well-known strength of neural networks is its ability to incorporate nonlinearities into the solution. Nonlinearity is common phenomena in the geosciences. Two other essential features, adaptability and generalization, are discussed below.

2.1. ADAPTABILITY

With supervised learning, the neural network learns to solve problems by example rather than by being programmed to follow a set of heuristics or theoretical mechanisms. For many real-world problems, precise solutions do not exist. In such cases, acquiring knowledge by example may be the only solution. In other words, when it is not possible to describe the logic of a problem or to predict behavior with analytical or numerical solutions to governing equations, traditional predictive analysis is difficult. Neural network analysis, however,

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

does not rely on prescribed relation, but rather seeks its own relation and thus may have an advantage over traditional predictive analysis.

It may be helpful to think of a neural network as a nonparametric, nonlinear, regression technique. In traditional regression techniques of modeling, one must decide *a priori* on a model to which the data will be fitted. The neural network approach is not as restrictive because the data will be fitted with the best combination of nonlinear or linear functions as necessary, without having to rigidly preselect the form of these functions.

2.2. GENERALIZATION

Generalization is the ability to abstract or to respond appropriately to data different from those involved in the training of the network. In learning to solve a problem, the network must encode the vital components of the problem into the weights. One might think of the weights on which the mature network has converged after "learning" the training examples as a distillation of the problem to these major components. Not only is a *well-trained* neural network skilled in generalizing the information held in these weights to new patterns it may be asked to judge, but the weights also form a storehouse of knowledge that may be explored or mined for insights about the problem as a whole.

Given a finite number of examples, the simplest network, the one with the fewest weights, which satisfies all input-output relations given by the training set, might have the best generalization properties. With too many weights the problem is similar to *overfitting* in regression analysis. This discussion is made more quantitative by defining the examples-to-weight ratio (EWR). The more examples with respect to the number of weights, the better the generalization properties of the network. In typical geophysical applications, we find that a rule of thumb for EWR is

$$\text{EWR} \geq 10. \quad (1)$$

Because fewer units in a network mean fewer weights, the EWR is maximized by minimizing the number of units in the network. It is important to note that the number of units from the input layer to the next layer is often most responsible for the large number of weights in a network. Hence, reducing the number of input features can be an effective way of reducing the number of weights. However, because small training-set learning is faster on a larger network, it is easy to fall into the trap of using large networks. In summary, just because a network has memorized a few examples perfectly does not mean that the network will perform well on new data; it is essential to consider the EWR discussed above.

3. Neural Network Models

There are numerous types of neural networks and they are discussed in many texts [1, 2, 3]. Following are three networks that have been particularly useful in seismic discrimination applications.

3.1. BACKPROPAGATION NETWORKS

We used backpropagation algorithms for seismic discrimination and yield estimation. An example architecture of a backpropagation network for seismic discrimination is shown in Figure 1. This network is a classic example of a supervised learning network and applies a learning algorithm, such that the network learns to associate the inputs with the corresponding outputs for all or

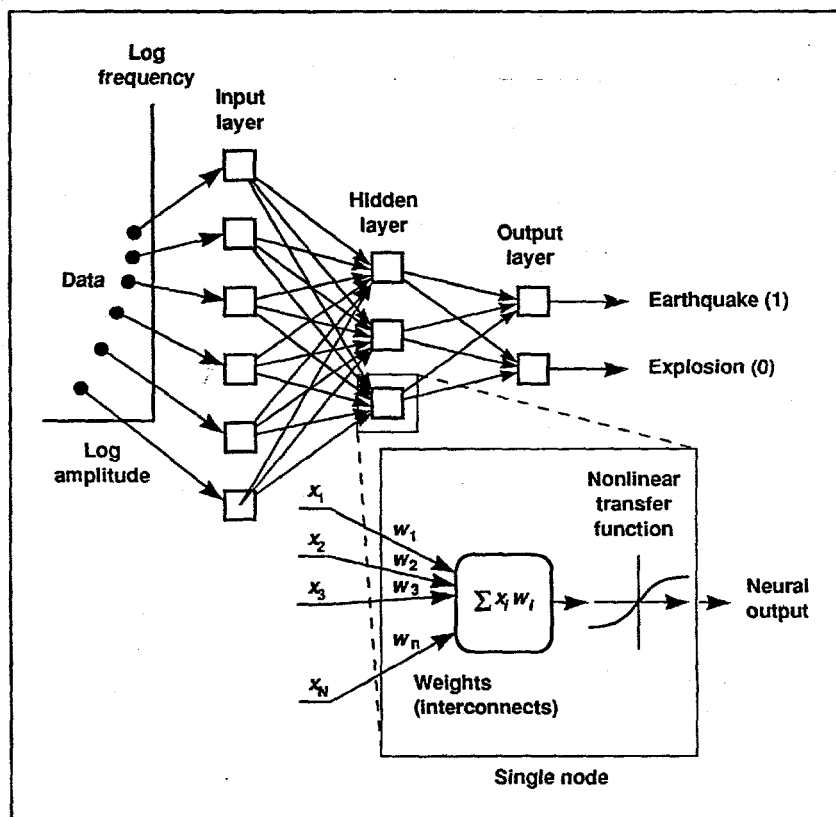


Figure 1. Architecture of the backpropagation neural network that used spectral inputs to discriminate seismic events.

most of the examples in a training set. *Backpropagation* is really a learning algorithm, and the network is a *multilayered feedforward* architecture. The backpropagation algorithm uses a gradient descent method to systematically modify the weights in the network so as to minimize the network output error. Once the network can solve a problem with a very large set of examples, the assumption is that the network can generalize from these examples to previously unseen data or test data.

3.2. SELF-ORGANIZING NETWORKS

Figure 2 depicts a self-organizing neural network (SONN) system that was used for seismic event classification. (When the number of classes is just two we use the term *discrimination*; otherwise we use the term *classification*.) In many problems it is useful to have a system that forms its own classification of the data from the training examples. In these types of problems, unsupervised learning networks are applicable. Unlike a backpropagation learning network, where the training examples must be explicitly specified at the output, an unsupervised SONN [4] clusters the data into groups that have common features. Hence, the SONN is often called a feature-mapping network.

Although the SONN can be viewed as a clustering algorithm, the clustering it forms is quite special. The network not only groups the input patterns into different clusters; it also organizes the clusters in a one- or multi-dimensional space according to similarity of the cluster centroids themselves. The motivation of the method is drawn from the knowledge that the brain utilizes spatial mappings; that is, different sensory and cognitive tasks are localized in different regions of the brain, and these regions have a definite topological order [4]. A supervised version of the SONN is the learning vector quantization (LVQ) network. This network can be used in discrimination problems.

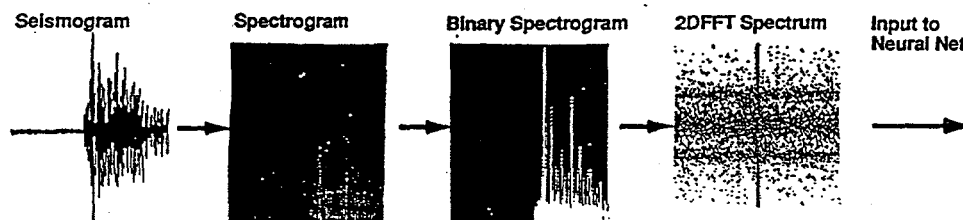


Figure 2. The self-organizing neural network system that was used to classify seven different types of events using binary spectrogram as input patterns.

3.3. RADIAL BASIS FUNCTION NETWORKS

We are developing radial basis networks to attack the depth estimation problem. Radial basis function (RBF) networks for the 2-D space are discussed here. The generalization of the algorithm to higher dimensions is straightforward. The basic step is to define a function of the form

$$f(x, y) = \sum_{j=1}^N \lambda_j \exp\left\{-\frac{[(x_j - x)^2 + (y_j - y)^2]}{2\sigma_j^2}\right\}. \quad (2)$$

Using the training set, the radial basis method solves the equations

$$f(x_k, y_k) = \sum_{j=1}^N \lambda_j \exp\left\{-\frac{[(x_j - x_k)^2 + (y_j - y_k)^2]}{2\sigma_j^2}\right\}, \quad (3)$$

where the 2-D function $f(x, y)$ is known at (x_s, y_s) and the objective is to determine the weights λ_j , the cluster centers (x_j, y_j) , and the cluster widths σ_j , while minimizing an error function. In matrix notation, equation (3) can be written as

$$\begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} f(x_1, y_1) \\ f(x_2, y_2) \\ \vdots \\ f(x_N, y_N) \end{bmatrix} \quad (4)$$

or

$$\Phi \underline{\lambda} = \underline{f}, \quad (5)$$

where

$$\phi_{ij} = \exp\left\{-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma_j^2}\right\}. \quad (6)$$

Note that the basis functions defined in equation (3) are radially symmetric distance functions, hence the name radial basis function. Other basis functions can also be used. The radial basis function defined in equation (6) is particularly attractive because the matrix Φ can be shown to be positive, definite [5], and therefore invertible. The weights of the radial basis network are obtained by inverting a set of linear equations once the width of the basis functions defined by the parameters σ_i have been estimated. Note, however, that the network is still nonlinear. Since the dimensions of the matrix Φ are proportional to the number of training examples, the matrix can become very large in many problems. Hence, a clustering algorithm is usually employed first, and these cluster centers are then used in evaluating equation (4).

4. Signal Representation

An appropriate representation of the data to the network is often the key to a successful design of an network. Three invaluable preprocessing techniques, Fourier and wavelet signal decomposition and projection methods in reducing the dimensions of the input, are briefly discussed next.

4.1. DECOMPOSITION

4.1.1. *Fourier Representation*

The transient nature of the frequency-dependent seismogram makes the spectrogram a powerful representation of the signal. We have found the time-frequency distribution to be a very useful in seismic event classification. In particular, the high-resolution maximum entropy spectrogram computed by segmenting the seismogram into small segments of overlapping time-series brings out features that helps the network to discriminate seismic events.

4.1.2. *Wavelet Representation*

The use of the recently popularized wavelet transform defined by

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{t=-\infty}^{\infty} x(t) h((t-b)/a) dt, \quad (7)$$

where $h(t)$ is the analysis wavelet, might prove to be particularly useful in seismic phase identification and depth estimation problems. The use of this representation is discussed in section 4, and an example of decomposition of a seismogram by wavelet transform is shown in Figure 3.

4.2. PROJECTION

Just as the decomposition techniques discussed above help in bringing out the vital features, the use of signal projection operation is important in reducing the representation to just these vital features. Often transformation of a signal might result in an increase in the dimensions of the input. This in turn results in a large number of weights. As stated earlier, an increase in the number of weights in the network is undesirable. Hence the use of projection operations such as the method of principal components or Hough transforms can used to reduce the input features.

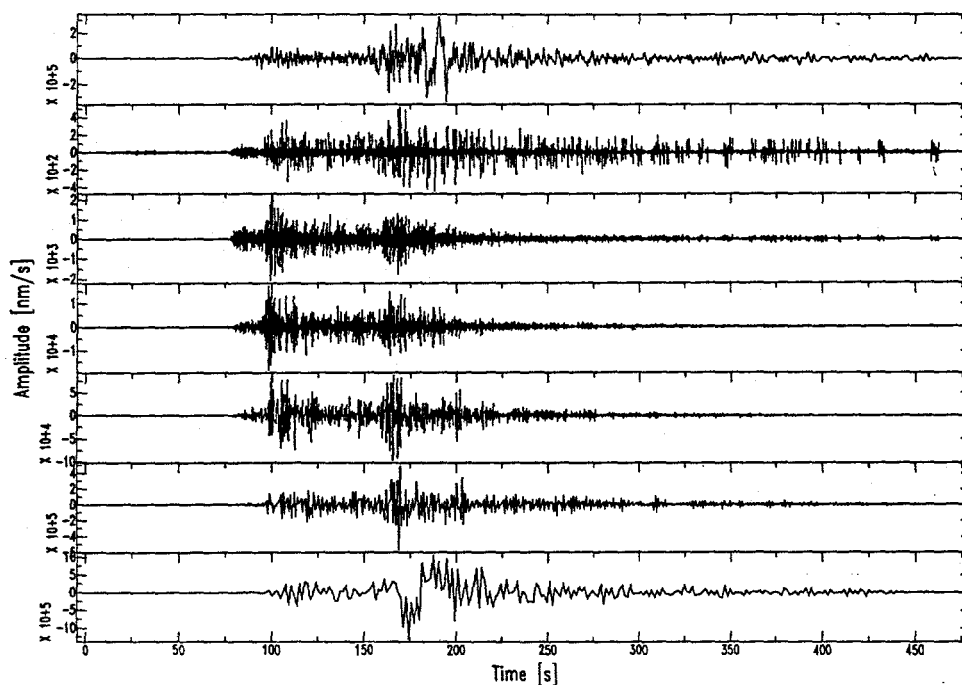


Figure 3. The wavelet decomposition of a seismogram into different scales. The top waveform is the recorded signal and the bottom waveforms are increasingly longer scale representations of the original data.

5. Application to Treaty Verification Problems

We have explored the use of neural networks on a variety of problems related to treaty verification. In this section we summarize these applications.

5.1. SEISMIC DISCRIMINATION

A neural network that performed quite well in distinguishing natural earthquakes from underground nuclear explosions is depicted in Figure 1. In a study conducted at four different stations, we used 83 western U.S. earthquakes and 87 Nevada Test Site explosions. The distance-corrected spectra of the Pn, Pg, and Lg phases were used as inputs. Details of the preprocessing and network properties of the network are summarized in [6]. For this data set, the rate of

TABLE 1. Performance of the neural network in discrimination between earthquakes (Q) and underground nuclear explosions (X).

	Number of events (Q/X)	Correct identification (%)	Misidentification (%)	Undecided (%)
Elko	80 / 79	97.5 / 97.5	0.0 / 2.5	2.5 / 0.0
Kanab	86 / 83	96.5 / 96.4	0.0 / 1.2	3.5 / 2.4
Landers	86 / 94	93.0 / 95.7	3.5 / 2.1	3.5 / 2.2
Mina	76 / 95	93.4 / 93.7	5.3 / 4.2	1.3 / 2.1
Four-Station Network	79 / 83	100. / 97.6	0.0 / 2.4	0.0 / 0.0

correct recognition for untrained data is over 93% for both earthquakes and explosions at any single station. Using a majority voting scheme on the four stations, the rate of correct recognition is about 97%. These results are summarized in Table 1.

We have recently developed a neural network for discrimination using SONN and LVQ methods on spectrograms. Others [7] have also developed networks for discrimination. Due to space limitations these methods are not discussed here; instead, the use of SONN on spectrograms for event interpretation is described.

5.2. SEISMIC EVENT CLASSIFICATION

Given the detection of a seismogram, we compute a time-frequency representation, the spectrogram, that describes the energy in the signal segment as a function of time and frequency. The mechanics of the computation of a spectrogram is quite straight-forward. Suppose we need to compute the spectrogram of a 200-s segment of seismic data. (If the data are sampled at 40 Hz, such a segment would consist of 8000 points). The computation of the spectrogram consists of breaking the segment into many sub-windows, called analysis windows, each with a length of 3 s, for example. The Fourier transform magnitude spectrum of the analysis windows is then computed. Adjacent analysis windows are usually overlapped by 67%, or 2 s in this example. Hence, there will be 300 analysis windows in the time domain for the 200-s segment seismogram. If we use a fast Fourier transform (FFT) of size 256, the number of nonredundant frequency points in the spectrum is 128. Therefore, the spectrogram can be viewed as a matrix with dimensions 128×300 , the rows representing frequency and the columns representing time.

We also developed a variation of this technique by computing what we refer to as the "binary spectrogram." The computation of the binary spectrogram consists of a thresholding operation in which the spectrogram values are made either 0 or 1 after comparing each value of the spectrogram to a threshold level obtained from the histogram of the signal and noise

TABLE 2: Performance of self-organizing neural network in event classification for seven types of events by class versus type of event (estimated probability).

Event classification	Car	Noise	Local	Regional	Teleseism
Class 1	0 (0.00)	1 (0.13)	5 (0.03)	2 (0.25)	0 (0.00)
Class 2	4 (1.0)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Class 3	1 (0.14)	4 (0.57)	2 (0.25)	0 (0.00)	0 (0.00)
Class 4	0 (0.00)	0 (0.00)	1 (0.33)	1 (0.33)	1 (0.33)
Class 5	0 (0.00)	0 (0.00)	0 (0.00)	2 (0.29)	5 (0.71)
Class 6	0 (0.00)	0 (0.00)	0 (0.00)	5 (0.71)	2 (0.29)
Class 7	4 (0.57)	1 (0.14)	1 (0.14)	0 (0.00)	1 (0.14)

characteristics of the data. The noise characteristic is measured on the data just before the arrival of the signal. The binary spectrogram helps in bringing out certain features more prominently and is a useful feature extraction step in this problem.

The training set for the SONN in one study [8] used 43 events. The events were labeled by a seismic analyst as automobiles (9), noise (6), local seismic (9), regional seismic (10), and teleseismic (9). Table 2 shows how these analyst-classified events are distributed throughout the seven clusters formed by the Kohonen model. Various numbers of clusters were tested; the cluster grouping results were best with seven clusters.

From the performance of the system on a limited number (43) of events, we conclude that the system is able to perform quite well with the fuzzy or soft event interpretation scheme for the different classes of events. However, to make the system even more portable and reliable, we need to develop the system using data from different regions of the world and integrate the knowledge base of those geologies into our system.

5.3. YIELD ESTIMATION

5.3.1. Seismic and Geophysical Parameters

The goal in the design of neural networks for yield estimation is to optimally combine different measurements to produce a better estimate. Because the output of the backpropagation network is usually the nonlinear sigmoid function, the network was modified such that the output neuron has a linear transfer function [9].

A number of factors can complicate the seismic yield estimation procedure and introduce uncertainties in yield estimates. For example, focusing and defocusing and pP interference effects add scatter to the time-domain magnitude measurements. Additionally, significant systematic variations in magnitude can be observed for explosions of a given yield in different media.

For a specific test site, these systematic effects can be partially attributed to coupling variations that are a function of the near-source geology. One of the geophysical parameters having the largest effects on seismic amplitudes is the gas-filled porosity, which is defined as the percentage of the material volume that is taken up by gas-filled voids. During expansion of the explosion shock wave, appreciable amounts of energy are absorbed during the crushing of gas-filled voids, causing a reduction in seismic amplitudes. Aside from the yield of the explosions, other factors affecting seismic amplitudes are the rock type, material strength, and depth of burial. The shape of the radiated seismic spectra from explosions is also known to be a function of emplacement conditions and yield. Examples of backpropagation networks with spectral and geophysical parameters as inputs are shown in Figure 4. Note that all these measurements are rather noisy. For these reasons seismic methods are less accurate than hydrodynamic methods in yield estimation. We discuss the hydrodynamic yield estimation neural network in detail in the next section.

5.3.2. Hydrodynamic Method with CORRTEX Data

Hydrodynamic theory allows us to use the speed of a shock wavefront to determine the yield of an explosion. On the basis of this theory, we developed a neural network to estimate yield of underground explosions from the shock wave radius-versus-time (RVT) data, as measured by continuous reflectometry for radius-versus-time experiments (CORRTEX). The proposed method replaces the subjective elements of conventional algorithms and produces significantly improved yield estimates. The network was trained with real hydrodynamic data, and its performance on unseen test events was studied. A backpropagation network was employed; using an architecture of 10 input units, a hidden layer with 11 hidden units, and 1 output unit (Figure 5). The network was trained

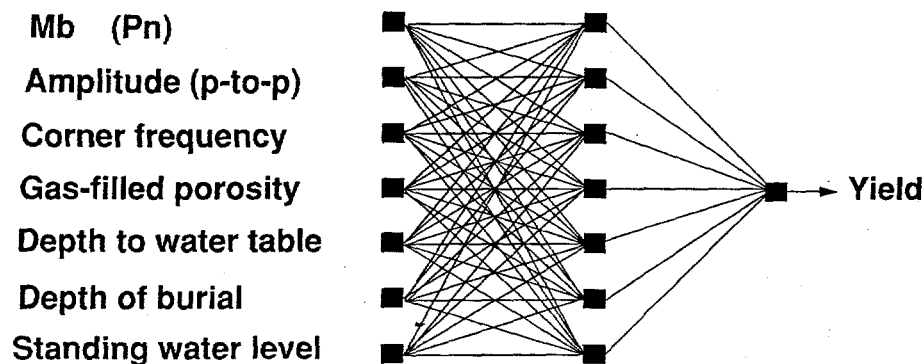


Figure 4. A backpropagation network was used to combine both seismic and geophysical parameters to obtain improved yield estimates.

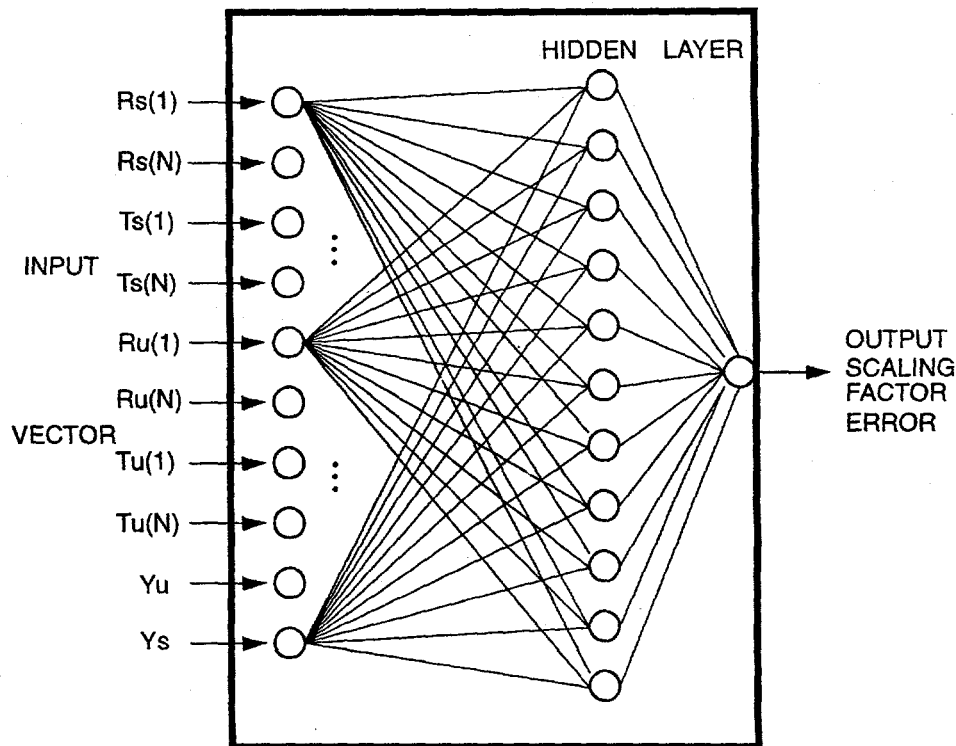


Figure 5. The architecture of the neural network that was used in the hydrodynamic yield estimation problem. The network consisted of ten input units, a hidden layer with eleven units, and a single output unit. The output represents the scaling factor error.

with thousands of input-output measurement vectors, the feasible input set, derived from the RVT data from only four other training or standard events (selected on the basis of the given RVT data from the unknown event). The feasible input vectors were propagated through the trained network, and the network output was used to select the optimum yield estimate. Elements of the input vector were center-of-energy (COE) offsets, shock-front radii, and time onset and interval of analysis for both the standard and the unknown events. The performance of the system is impressive. For 24 tuff events, at the 95% confidence interval level, the F value was only 1.17, compared with 1.35 for an existing expert system measuring the same events. The study is described in [10].

5.4. SEISMIC PHASE IDENTIFICATION AND DEPTH ESTIMATION

Identifying the regional seismic phases and estimating depths of a seismic source are two of the most challenging and important tasks in seismic discrimination. Preliminary studies [11, 12] report that neural networks might be useful in solving these problems. We are developing networks for depth estimation from real seismograms. Although our results are preliminary, we are having modest success in depth estimation using wavelet decomposition followed by a radial basis network.

6. Summary

Neural networks are useful tools in solving many different types of problems. However, the choice of the correct network and appropriate preprocessing is often the key to developing a powerful network. Another important element of the design process is a large training set, both for training and for testing.

7. References

1. Beale, R., and Jackson, T. (1990) *Neural Computing: An Introduction*, IOP Publishing Ltd., Bristol.
2. Hertz, J., Krogh, A., and Palmer, R. G. (1991) *Introduction to the theory of neural computation*, Addison-Wesley Publishing Company, MA.
3. Simpson, P. K. (1990) *Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations*, Pergamon Press, New York.
4. Kohonen, T. (1989) *Self-Organization and Associative Memory*, 3rd ed., Springer-Verlag, Berlin.
5. Powell, M. J. D. (1987) Radial basis functions for multivariable interpolation: a review, *Algorithms for Approximation*, J. C. Mason and M. G. Cox, eds., Oxford University Press, Oxford, 143-167.
6. Dowla, F. U., Taylor, S. R., and Anderson, R. W. (1990) Seismic discrimination with artificial neural networks: preliminary results with regional spectral data, *Bull. Seis. Soc. Am.* 80, 1346-1373.
7. Cercone, J. M., and Martin, J. R. (1994) *An Application of Neural Networks to Seismic Signal Discrimination*, Phillips Laboratory, Hanscom Air Force Base, Massachusetts, Scientific report No. 3, PL-TR-94-2178.
8. Maurer, W. J., Dowla, F. U., and Jarpe, S. P. (1992) Seismic event interpretation using self-organizing neural network, *Proceedings of the SPIE*, vol. 1709, 950-958.
9. Leach, R. L., Dowla, F. U., and Vergino, E. S. (1993) Yield estimation using bandpassed-filtered seismograms, *Bull. Seis. Soc. Am.* 83, 488-508.

10. Dowla, F. U., Leach, R. L., Glenn, L., Moran, B., and Heinle, R. (1993). Neural network methods in hydrodynamic yield estimation, *Pure and Applied Geophysics* 140, 427-454.
11. Lacoss, R. T., Cunningham, R. K., Curtis, S. R., and Seibert, M. C. (1991) Artificial Neural Networks for Seismic Data Interpretation, Lincoln Laboratory, Lexington, Massachusetts, Technical Report ESD-TR-91-170.
12. Perry, J. L., and Baumgardt, D. R. (1991) Lg depth estimation and ripple fire characterization using artificial neural networks, *Neural Information Processing Systems*, Lippman, Moody, and Touretzky, eds., Morgan Kaufmann Publishers, San Mateo, California, 544-550.

Acknowledgments

The work described in this paper is a result of collaboration with many colleagues at LLNL. In particular, I would like to acknowledge the valuable collaboration of Lew Glenn, Eric Johanssen, Richard Leach, William Maurer, Leah Rogers, and Steve Taylor over the past four years in various neural networks problems. I would also like to thank Sue Stull for her editorial help. Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-Eng-48.