Can Neural Networks be used in Data-Poor Situations?

William Silvert
Department of Fisheries and Oceans, Habitat Ecology Section
Bedford Institute of Oceanography, PO Box 1006
Dartmouth, N.S., CANADA B2Y 4A2
Silvert@scotia.dfo.ca

and

Martin Baptist
WL | Delft Hydraulics
Marine and Coastal Management
P.O. Box 177
2600 MH Delft, Netherlands
Martin.Baptist@wldelft.nl

Abstract

A major limitation on the use of neural networks is that they require large amounts of data for training. This is a serious problem in most ecological applications, because data are usually very scarce. It is therefore essential to find ways of making the training of neural networks more efficient so that they can be used in situations where the number of data is small. We explore several ways of doing this, including constraining the system on the basis of prior information and the use of transformed variables to reduce the effective number of degrees of freedom. The more successfully we can integrate neural networks with other analytical procedures based on scientific understanding of how ecosystems function, the more effective neural networks will be in dealing with ecosystems for which only limited data are available.

1. Introduction

Perhaps the greatest problem that is faced in most attempts to use artificial neural networks for ecological applications is that the quantity of data is often very limited. Although there are a few cases where large amounts of data are available, as in the case of remote sensing or observations based on automatic telemetry, it is far more common to have to deal with limited and irregularly spaced data, and the data may not always be strictly comparable due to variations in environmental conditions between sampling periods. In most situations the collection of field data is both time-consuming and expensive. Since the training and testing of neural networks is very data-intensive, this poses serious obstacles to the development of neural network applications in ecology.

The field of marine benthic ecology is one in which data are difficult to obtain, and large databases are almost unknown. Most benthic data are obtained by laborious analysis of individual samples, usually cores or grabs, although automated sediment traps and other computerised instrumentation are beginning to be used. Because of these data restrictions, the prospects for using artificial neural networks in the analysis of benthic data are not promising. On the other hand, analysis of benthic data often requires highly specialised expertise that is not commonly available, so the incentive to use artificial neural networks is strong.

The material in this paper is based on an effort to facilitate the analysis of geochemical cores by developing an artificial neural network with a very limited data set. After careful review of the data and elimination of unreliable samples, we were left with a very small number of cores, and the prospects for development of a neural network seemed very dubious. Still, we feel that we made substantial progress by suitable preprocessing of the data, and we feel that this approach may have applicability to other applications in ecology and similar data-poor fields.

2. Neural network training and its limitations

The most common neural networks are feedforward neural networks that are trained using error back propagation. This is a training method in which the network is supplied with input values and also with the desired output values. The weights in the network are adjusted based upon the error between the expected output and the computed network output until this error is minimised. In a reasonably complex network the number of weights is large and unless there are many data pairs, the minimisation process may not give meaningful results. For this reason, the application of neural networks to ecological data does not always lead to reliable models.

As with most empirical (e.g., statistical) modelling approaches, one divides the total data set into two parts, one of which is used for fitting the model (which is referred to as "training" in the terminology of artificial neural networks), and the other of which is used for testing it. When the data set is very large, typically involving thousands or even tens of thousands of data pairs, this is relatively straightforward. With smaller data sets it can be difficult to decide how much of the data can be used for training the model while still leaving enough data to test it adequately. One approach is to repeat the process several times, selecting a training set at random, then fitting the model and testing it with the remaining data.

One must however be careful not to "over-train" the network, meaning that after too many iterations one stumbles upon a model which fits the training data almost perfectly, without actually providing a good representation of reality. If the test set is too small, the network may give adequate test results, leading to acceptance of a model which is not very good. This is especially true if the model has too many degrees of freedom, which is the result of including too many hidden neurons – this is analogous to the statistical problem of fitting a model with too many parameters, such as a polynomial of too high order.

Another issue related to the overtraining problem is the bias-variance dilemma (Geman et al. 1992). It can be demonstrated that the mean square value of the estimation error between the function to be modelled and the neural network consists of the sum of the (squared) bias and variance. With a neural network using a training set of fixed size, a small bias can only be achieved with a large variance (Haykin 1994). This dilemma can be circumvented if the training set is made very large, but if the total amount of data is limited, this may not be possible.

As pointed out in the introduction, ecological data sets are usually small. Unfortunately ecological processes are often non-linear and poorly understood, and the limited data sets are barely adequate. Since the processes may not be adequately understood, the use of neural networks to generate empirical models of complex behaviour seems a promising technique for a better replication and understanding of ecosystem behaviour. On the other hand, the small size of the data sets make it very difficult to train and test these models well enough to have confidence in the results.

The situation is not as hopeless as it may appear. The field of artificial neural networks is developing very rapidly, and new approaches resolve many formerly intractable problems. For example, general regression neural networks (Specht 1991) offer a promising way to deal with small data sets, and have been used successfully to construct a neural network for a data set not much larger than the one used in this project (Megrey 1999). There are some other neural network configurations that are capable of handling smaller data-sets, such as Kohonen networks, but the use of sophisticated second-generation approaches is not the issue we are trying to address in this paper. Instead we have tried to explore ways of dealing with small data sets when one uses the most popular configuration of neural networks, feedforward error back propagation networks.

When only a relative small data set is available for the application of a neural network, a number of drawbacks arise. Firstly, one has to split the already small data set into a training set and a testing set. Secondly, overtraining is more likely to occur with small datasets, because the degrees of freedom in the network rapidly increase with the number of neurons. For a good model, the number of data pairs should exceed the number of weights in the neural network.

3. The Geochemical Data Problem

Geochemical profiles sampled underneath fish farms provide valuable data on the benthic impacts of these farms, but the interpretation of these data is a complex process requiring scientific sophistication and understanding of benthic processes. It is difficult to avoid a degree of ambiguity and subjectivity in the interpretations, although this is a more general problem in the analysis of scientific data than is commonly admitted (Silvert 1997). Some typical geochemical profiles are shown in Figure 1. These types of profile are commonly found for organic carbon and organic matter, and, in the case of hyperenriched sediments, porewater concentrations of nutrients and hydrogen sulphide. The first profile, Fig. 1a, shows a continuous decrease in carbon levels presumably reflecting constant deposition and degradation. The second profile, Fig. 1b, shows a flat plateau suggesting that the sediments have been mixed by bioturbation. Since bioturbation is indicative of a viable benthic community, the second of these profiles would normally be interpreted as showing less of an environmental impact than the first.

Angel et al. (1998) used data from dive logs to classify environmental impacts of a fish farm in the Red Sea in terms of four fuzzy sets; Nil, Moderate, Severe and Extreme impact. This paper used only visual data recorded by divers, but on some of the sampling dates the divers also took sediment cores, and these can also be used as indicators of environmental impact. The number of useful cores is substantially less than the number of dive logs, but because it is widely felt that chemical analysis of such cores provides a more objective measure of environmental impact than the dive logs, which involve subjective elements, research is continuing on the use of these cores as environmental indicators. In particular, sediment profiles for loss on ignition (LOI) data are being investigated. LOI is a measure of carbon content, and since carbon deposition is one of the major forms of environmental effect from fish farms (due to faecal settlement and waste feed), it is a critical indicator of both current and past deposition rates, as well as of bioturbation.

Although preliminary results indicate good agreement between the classifications based on the dive logs and those obtained from the LOI data, the analysis of LOI data is extremely difficult and requires a great deal of expertise. This makes it a prime candidate for neural network analysis, since the value of LOI data for monitoring purposes would be greatly enhanced if an expert system could be used for the analysis, and there did not seem to be any reliable way of developing a rule-based system.

We subsequently undertook to develop an artificial neural network to simulate the impact-assignment process in an effort to develop an expert system that does the same analysis, as described in Baptist et al. (1998).

The sediment profiles contain eleven sampling depths. The values for loss on ignition (LOI) at each depth were used for the input neurons, so there were eleven input neurons. The fuzzy membership values for the four impact classes were used as outputs – these memberships sum to one, so there were three independent output neurons. The best results with the feedforward error back propagation neural network were obtained with a hidden layer of eleven neurons. Configurations with fewer hidden neurons, even down to three, were also tried. Although the fit to the training set was reasonable, the fit to the test set worsened with a decreasing number of hidden neurons. The total network therefore had 165 connections. The dataset of LOI profiles measured over a four-year period was small and contained only nineteen suitable profiles. Baptist et al. (1998) were faced with a typical problem when using neural networks for ecological problems: the dataset was too small in relation to the number of degrees of freedom in the neural network. Their results were typical for this kind of problem, the training set (fourteen profiles) showed an almost perfect fit, whereas the test set (five profiles) had considerable discrepancies. Figure 2 gives a sample of the results that were obtained by applying the artificial neural network to the scores for “Moderate-impact”. Even though the test results were still acceptable, this cannot be considered very significant in light of the small amount of data available for the test procedure.

4. Preprocessing data

One of the reasons why neural networks are inefficient is that they are usually trained with raw data that ignores the understanding and insight that a human expert can bring to bear. For example, in the classic pattern recognition problem of training a system to recognise human faces, it is considered an achievement when an artificial neural network “learns” to ignore whether the subject is wearing eyeglasses. Edelman et al. (1998) discuss the differences one encounters training a neural network to recognise faces when the full face is visible and when the hair is covered by a bathing cap. But we all know that eyeglasses are almost useless for identification purposes, and hair shape and colour can change rapidly. If we could teach the neural network in advance to ignore eyeglasses and hair style, and to concentrate on facial structure, it should be possible to make the training process much more efficient.

One way to interact with the training process is to preprocess the data before it is fed into a neural network. Preprocessing has been applied to pattern recognition problems, although more for the purpose of simplifying the vast amount of data contained in a matrix of pixel values than for the present purpose of extracting information as efficiently as possible from a limited data set (Huntsberger et al. 1998). We might process the images through a graphics program that would erase eyeglasses and facial hair, and generate just a three-dimensional profile that reflected the underlying bone structure. We are not aware of any effort to investigate this approach, although the underlying idea of looking for basic geometric structure has been discussed by Edelman (1998). We did however look at ways of transforming the data for the LOI profiles in such a way as to simplify the work of the neural network, and we feel that the approach led to promising results.

When an expert on benthic ecology and geochemistry assigns classifications based on different LOI profiles, he makes use of subjective factors even though the data themselves are objective numerical values. An expert uses his knowledge of natural processes to analyse the shape and overall magnitude of the overall profile or parts of the profile (for example the upper 5 centimetres) to arrive at an interpretation of the degree of impact. However, since a neural network only sees the numbers in the profile as independent variables, it doesn't even have the valuable information that the depths of the samples are contiguous. It may therefore improve the neural network performance if the data are preprocessed and expert-based discriminants are used as input to the neural network. This can also simplify the neural network layout, speed the training process and result into a network with a smaller number of degrees of freedom.

One way in which we can link this type of human pattern recognition with artificial neural networks is to transform the data using orthogonal functions. For example, Edelman et al. (1998) use principal component analysis to represent the spatial data constituting a pixel image of a face in terms of a small number of eigenfunctions (i.e., orthogonal functions) which are then used as input variables. This procedure seems relevant to the problem of analysing the LOI profiles, except that we have to keep in mind that each of these profiles contains at most eleven points, while the facial images consisted of over 15,000 pixels. The use of data transformations in pattern analysis is more a way of reducing unmanageably large data bases than a way of using small data sets efficiently. Even so, the use of orthogonal transformations to describe the connectivity of input data points seems promising in both situations.

Baptist et al. (1998) experimented with ways of transforming the data with orthogonal functions, such as Fourier transforms, to generate input variables that more closely correspond to the patterns that experts see when they analyse these kinds of data. Fourier transforms distinguish particular patterns, and at least for the lower-frequency transforms, each Fourier component emphasises a specific part of the profile and might serve as a discriminant. For example, the lowest frequency components describe the total carbon content and the gradient of the profile. Since Fourier components are based on sine and cosine functions, they have a close correspondence to the wavelets used by Huntsberger et al. (1998). The first five Fourier components were chosen as input neurons. This way, each Fourier component distinguishes a particular pattern in the LOI profiles. Consequently the number of input neurons were reduced to five in a network with five hidden neurons and four output neurons. This brings back the number of weights to 45. A sample of the results of this neural network for the "Moderate-impact" scores are shown in Figure 3 and were comparable with those of the raw LOI profiles. We found that the Fourier components were very sensitive to variations in the LOI values. Furthermore, the number of data-pairs was too small to develop a good neural network.

5. Results

A comprehensive discussion of the results of this study would be complicated by the fact that the expert assessments that were used to train and test the neural network were expressed in terms of fuzzy membership functions. Because of the additional complexity that this introduced to the analysis, we refer the reader to Baptist et al. (1998) for a full discussion of the theoretical background of the study and of the detailed results. In this paper we want to focus on the problems of the small data set and the issue of whether transforming the data added to the efficiency and effectiveness of the process. We feel that these results are strongly suggestive, but not conclusive.

The analysis of the data was carried out with both the raw input data (i.e., LOI values as a function of depth) and with transformed values, using the first four or five Fourier components only. We were surprised to see that in both cases the neural networks worked quite well – it is not surprising that they fit the training sets of course, since there were so few data and so many parameters, but the test sets were reasonable. Given the small size of the test sets we cannot attach very much significance to this result, but they do suggest that the use of neural networks for even very small data sets may prove practicable.

We were however also surprised to see that the use of a smaller number of transformed variables as inputs did not appreciably improve the fitting of the network. On more careful examination of the network we discovered that when the raw data were used, the output was dominated by the first input neuron, corresponding to the LOI value just below the surface (0.5 cm), so from this point of view the raw data was already in a form that might not be improved by transformation.

6. Discussion and conclusion

When data are scarce it might be useful to preprocess (transform) the data before they are used to train a neural network. The transformation technique must be chosen with care and should be based on ecological knowledge of the system. A preprocessing of data can reduce the number of input neurons and therefore reduce the problem of overtraining. It may also help to understand the process handling inside a neural network, making it less of a black box then it is when blindly applied to incoherent data sets.

However, preprocessing based on understanding requires that the understanding be reliable, and we don't always know that this is the case. Huntsberger et al. (1998) discuss the distinction between feature-based systems and image-based systems, and we have not been able to identify clearly the extent to which the experts who analysed the LOI profiles relied on features (such as the near-surface values) as opposed to evaluating the entire profile. They were also surprised to learn that the top part of the profile played such a major role in determining their evaluations, but in retrospect this may not be unreasonable even if we accept their assertions that they looked at the total shape of the profile. For example, a core that displays evidence of bioturbation is likely to have lower levels of surface carbon because it has been mixed deeper into the sediments. Although we feel that discussions with the experts can help resolve these issues, it is difficult when dealing with such a small data set as the one that was available for this study.

We are not convinced that Fourier transforms were the best choice of transform functions for this project, since they are best for analysing data that are distributed over a uniform interval. Given the shape of the LOI profiles it might have been preferable to use Bessel or Laplace transforms which offer a better representation of functions which are large near the origin and taper off as one moves away from it, but neither the amount of data nor the time available permitted exploration of this idea.

The use of preprocessing data to refine the inputs to artificial neural networks has been studied only in a few cases, mainly in the context of reducing large quantities of graphic data to manageable scale, and whether the same approach can be successfully applied to very small data sets is not clear. Even in the relatively advanced field of image processing and facial recognition, it has not been shown conclusively that preprocessing with orthogonal transformations is advantageous (Edelman et al. 1998). However we feel that the approach is promising and deserves further investigation. Certainly if neural network theory is to be applied to small and highly variable ecological data sets, the focus must be on how to use all the data as efficiently as possible, and we should not be using the data to establish patterns that are already known.

References

Angel, D., Krost, P., and Silvert, W., 1998. Describing benthic impacts of fish farming with fuzzy sets: theoretical background and analytical methods. J. Appl. Ichthyology **14:** 1-8.

Baptist, M.J., Silvert, W., Krost, P., and Angel, D. L., 1998. Assessing benthic impacts of fish farming with an expert system based on neural networks. Bull. Canadian Society of Theoretical Biology, Fall 1998. The CSTB Bulletin is an electronic journal accessible at URL **Error! Hyperlink reference not valid.**/, and the paper itself is located at URL **Error! Hyperlink reference not valid.**ff/silvert/bska/.

Edelman, B., 1998. Spanning the face space. J. Biol. Systems **6**: 265-280.

Edelman, B., Valentin, D., and Abdi, H., 1998. Sex classification of face areas: how well can a linear neural network predict human performance? J. Biol. Systems **6**: 219-240.

Geman, S., Bienenstock, E. and Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Computation **4**: 1-58.

Haykin, S., 1994. Neural Networks; A Comprehensive Foundation. Macmillan College Publishing Comp., New York, USA.

Huntsberger, T., Rose, J., and Ramaka, S., 1998. Fuzzy-face: a hybrid wavelet/fuzzy self-organizing feature map system for face processing. J. Biol. Systems **6**: 281-298.

Megrey, B. A., 1999. Using neural networks to examine relationships between features of the environment and recruitment variation in exploited marine fish stocks. [This volume???]

Silvert, W., 1997. Ecological impact classification with fuzzy sets. Ecological Modelling **96**: 1-10.

Specht, D., 1991. A general regression neural network. IEEE Trans. on neural networks **2:** 568-576.

Figure Captions

Figure 1. Two typical sediment profiles. The one on the left, 1a, shows simple deposition and degradation. The second, 1b, exhibits probable bioturbation.

Figure 2. Sample results obtained with an artificial neural network (ANN) fit to the raw “Moderate-impact” scores of the experts.

Figure 3. Sample results obtained with an artificial neural network (ANN) fit to the Fourier transforms of “Moderate-impact” scores of the experts using the first five Fourier components.

Figure[s]

Figure 1a

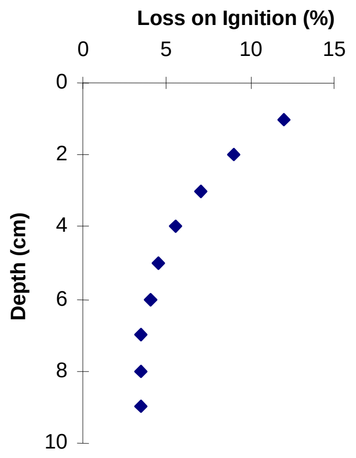
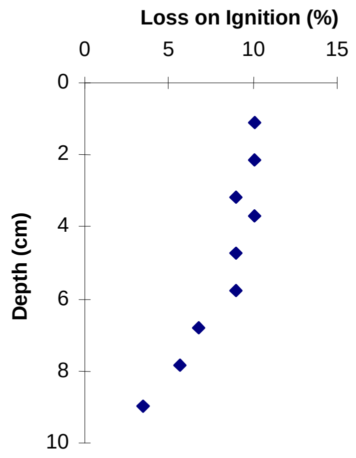
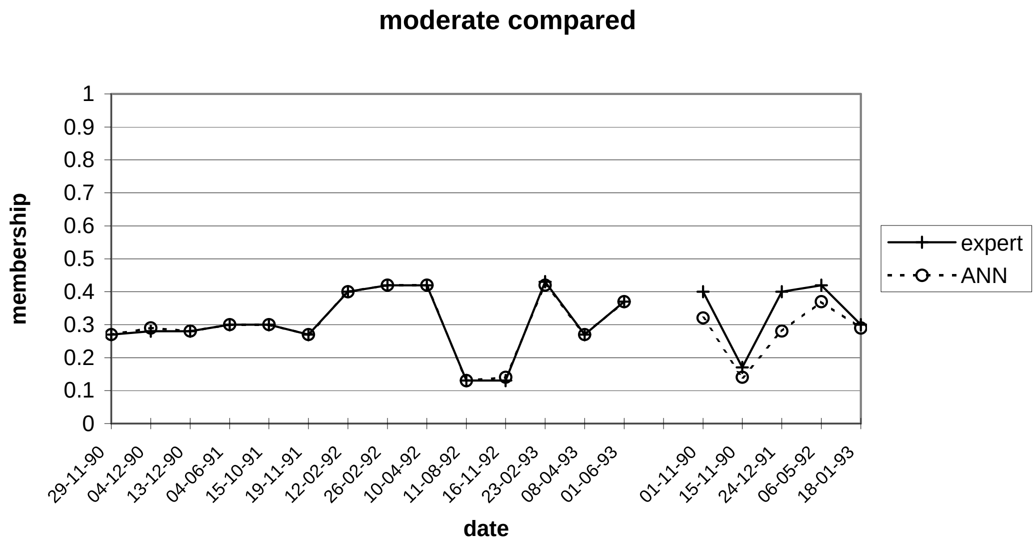
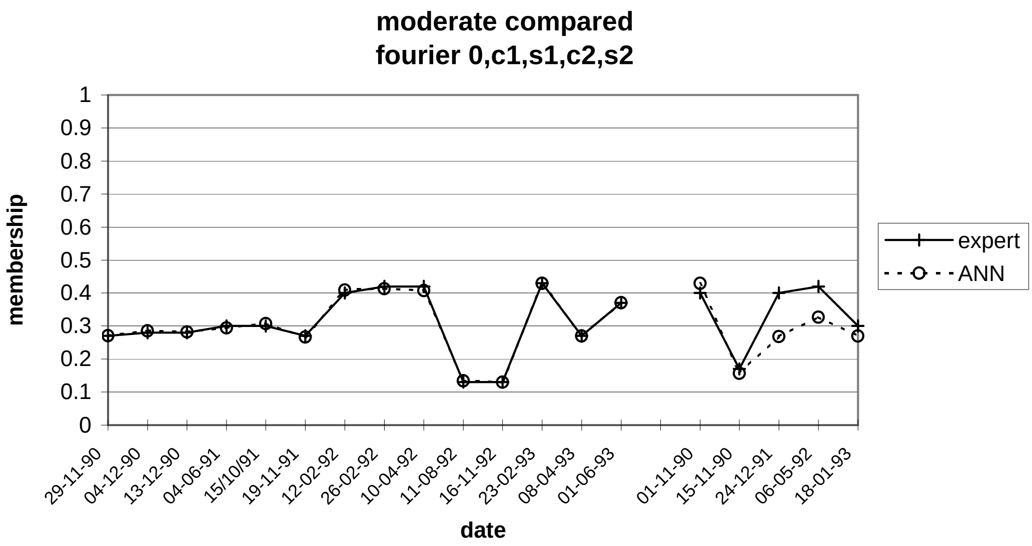


Figure 1b






moderate compared
fourier 0,c1,s1,c2,s2
membership
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
29-11-90
04-12-90
13-12-90
04-06-91
15/10/91
19-11-91
12-02-92
26-02-92
10-04-92
11-08-92
16-11-92
23-02-93
08-04-93
01-06-93
01-11-90
15-11-90
24-12-91
06-05-92
18-01-93
date
expert
ANN