International Symposium on
HYDROACOUSTICS AND ULTRASONICS
EAA Symposium (formerly 13th FASE Symposium)
Gdańsk-Jurata, 12-16 May 1997

Comparison of selected classifiers in a sea-bottom recognition task.

R. Komendarczyk

Acoustics Department, Technical University of Gdańsk
80 - 952 Gdańsk ul. Narutowicza 11/12 Poland

Abstract

This paper presents a test of three following classifiers: minimum-distance classifier, feed-forward neural network with backpropagation learning scheme and neuro-fuzzy classifier based on NEFCLASS architecture. They have been applied to the sea bottom type classification task over different input spaces. The experiment proved high efficiency of the minimum-distance classifier and the neural network, NEFCLASS performance had been rather poor. Generalization properties of those classifiers are also investigated. Additional conclusions concerning classifiers topology are presented.

Introduction

One of the possible approaches to the seabed classification task is to construct such a decision function (classifier) which could divide an input vectors space in such a manner that a correct generalization is obtained as a result. Usually, a classifier is built as an approximation of some prior knowledge which composes a set of training points. If a correct generalization is obtained, one can successfully use a classifier for identification of unknown data.

Generally in any seabed classification system classifiers are present, but usually their domain (input space) is a space of signal parameters. Recently, many researches [1],[2],[9] put a lot of effort into finding such a set of signal parameters which form the input space where bottom type classes are easily separable and a simple classifier can handle the task of recognition. However, it could happen that such a classifier is more susceptible to errors. Commonly it is known that they can be caused by the following features of the input space:

1. The input vectors may be inadequate for distinguishing between the different classes.
2. The input vectors may be highly correlated.
3. The decision boundary may have to be curved.
4. There may be distinct subclasses in the data.
5. The input space may just be too complex.

It is known that more advanced supervised classifiers (like Multi Layered Perceptrons (MLP), Neuro-Fuzzy Systems and Radial Basis functions RBF) can handle the classification task (even with the above limitations) with higher accuracy, since they can approximate any continuous mapping on R^n and have sufficient ability to generalize (extrapolate). The problem that arises most often when application of those systems is considered, is the choice of the appropriate classifier topology. Commonly the quantity called *VC(Vapnik-Chervonenkis) - dimension* [5] can be used to estimate (input space) separation ability for the given family of classifiers. However, the optimal topology for a given classification task is a tough problem, which can only be solved experimentally.

Review of classifiers tested in the experiment.

In this experiment generalization properties of three following classifiers have been studied: minimum-distance classifier, feedforward neural network with backpropagation learning scheme and neuro-fuzzy system NEFCLASS. The experiment was made over a set of 3714 bottom echo pulses, classified in four classes : *rock, sand, soft-sand, soft-mud.* A benchmark of selected classifiers over different input spaces is presented.

In a general case any classifier can be understood as a certain function:

$$F: X \to D$$

where:

X - a set of classified objects

D - a set of classes (frequently extended by an additional "unknown" class" which is assigned to a given vector x when the vector x cannot be explicitly classified into one of the remaining classes).

Usually the term supervised classifier denotes a classifier whose method of construction requires a certain set of pre-classified objects from space X, i.e. a certain sequence of pairs $Z=\{(x_i,d_i) \in X\times D;\ i=1..k\}$. The neural networks described below and the NEFCLASS system find the form of the function F through approximation so that:

$$\forall_{i=1...k} F(x_i) = d_i, \qquad (x_i, d_i) \in Z.$$

In the case of the minimum distance classifier function F is constructed directly on the basis of the training set Z.

Minimum - distance classifier.

This is one of the simplest classifiers. To classify an unknown input vector x is to assign this class $d_i \in D$ $((x_i,d_i) \in Z)$ for which the distance $\|x_i - x\|$ (the experiment assumes Euclidean distance) is shortest. A detailed scheme of this classification is the following:

1. For each pair $(x_i,d_i) \in Z$ the distance $\|x_i - x\|$ is calculated.
2. If there is exactly one vector $x_p \in X$ $((x_p,d_p) \in Z)$ for which $\|x_p - x\| = \min_{i=1...k}(\|x_i - x\|)$. Then x is assigned class d_p, that is $F(x)=d_p$.

3. In the case when there is more than one vector x_p for which the above condition of minimum distance is met, the classifier generates the decision: "unknown".

Feed-forward neural network with a back-propagation learning schema.

The algorithm of error backpropagation is the basis of current work on neural networks training. The algorithm tells us how to change the weights $w_{i,j}$ in any given feed-forward neural network that is supposed to learn the training set Z [6]. The general form of mapping which is the feed-forward neural network is given below:

$$F: R^n \to R^k \qquad F([x_1,\ldots,x_n]) = [y_1,\ldots,y_n]$$

where:

$$y_i = \sigma\left(\sum_{m=1}^{k_1} w_{L,m} h_{L-1,m}\right), \quad i = 1\ldots k$$

$$h_{L-1,i} = \sigma\left(\sum_{m=1}^{k_2} w_{L-1,m} h_{L-2,m}\right), \quad i = 1\ldots k_1$$

...

$$h_{1,i} = \sigma\left(\sum_{m=1}^{n} w_{L-1,i} x_m\right), \quad i = 1\ldots$$

n- dimension of the input vector

k- dimension of the output vector

$k_1 \ldots k_{L-1}$ - number of neurons in the particular hidden layers of neural networks

σ - function of neuron activation, usually in the form of a selected sigmoidal form:

$$\sigma(h) = \frac{1}{1+e^{-2\beta h}} \qquad or \quad \sigma(h) = \tanh(\beta h);$$

The weights $w_{i,j}$ are determined in the minimization process using the algorithm of gradient descent, a selected criterion function, usually it is [4]:

$$E(w) = \frac{1}{k}\sum_{l=1}^{N} \|F(x_l) - d_l\|,$$

where: $(x_l,d_l) \in Z$, N - a power of set Z.

w - vector of all weights $w_{i,j}$

$\|.\|$ - selected metric (differentable).

The above error measurement is a continuous and differentiable function of all weights, that is why we can apply the gradient decend algorithm. Modification of the weights at a random step of the iterative algorithm is given as [4]:

$$\Delta w_{i,j} = -\eta \frac{\partial E}{\partial w_{i,j}}$$

where :

η -constant higher than zero

NEFCLASS (NEuro-Fuzzy CLASSification) [7]

A neuro-fuzzy system is usually a fuzzy system that uses a learning algorithm which is derived from the neural network theory. The changes computed by the learning algorithm are based on local information only, and the changes are also carried out locally. The fuzzy system is usually viewed as a special 3-layer feed-forward network architecture, where the units of the second layer represent the fuzzy rules. The fuzzy sets are represented as *fuzzy weights* on the connections from the input to the hidden layer. NEFCLASS is used to derive fuzzy rules from a set of data that can be separated in different crisp classes. The fuzzy rules describing the data are of the form:

if x_1 is μ_1 **and** x_2 is μ_2 **and ... and** x_n is μ_v **then** the pattern $[x_1,\dots,x_n]$ belongs to class i, where $[\mu_1,\dots,\mu_n]$ are the fuzzy sets. The task of the NEFCLASS model is to discover these rules and to learn the shape of the membership functions to determine the correct class category of a given input pattern. The learning algorithm for NEFCLASS is described completely in [8]. In the following part we will only give a short overview:

- **Initialization:** For each feature there is an input unit, and for each class there is one output unit. For each input unit an initial fuzzy partitioning is specified (e.g. a number of equally distributed triangular membership functions).
- **Rule Learning:** NEFCLASS starts without rules, and inserts fuzzy rules into the system during a first run through the training data. In a second run the rules are evaluated, and only the best *r* rules are kept, where *r* is given by the user. It is also possible to keep the best rules class.
- **Fuzzy Set Learning :** For training of the membership function a backpropagation schema is used. Depending on the output error for each rule unit a decision is made, whether the activation value has to be higher or lower. Each rule unit then changes its membership functions by changing their support. The user can specify several constraints so that the changes of the membership functions do not change the semantics of the underlying fuzzy model.

The experiment

A general schema of echo signal envelope classifications can be defined as the following mapping assumptions:

$$C[a,b] \xrightarrow{i_p} R^n \xrightarrow{F_k} D$$

where:

$C[a,b]$ - space of continuous functions in the range [a,b] with values in R (includes in a natural way all the envelopes of signals reflected from the bottom)

i_p - task of mapping C[a,b] in R^n.

F_k - selected classifier

D= {rock, sand, sand, mud}.- set of classes of bottom types

The paper presents tests for three selected mappings i_p, and for the minimum-distance, NEFCLASS and neural networks classifiers described earlier. The classifiers were trained consecutively for ten training sets $B_{1,p}$ ... $B_{10,p}$ obtained from a basis of the given envelopes and were then tested on a set $B_{1p,}$ maximal in terms of its power. The percentage of correct decisions for each classifier has been illustrated in the diagram for each stage of the experiment. Sets $B_{j,p}$ were constructed as follows:.

1. From a set of 7428 envelopes of the first echo, subsets $\{A_1 \dots A_{10}\}$ were selected respectively of 3714 envelopes 2476, 1857, 1485, 1238, 1061, 928, 825, 742 and 372 envelopes. While sets $\{A_i\}_{i=1\dots10}$ did not constitute a declining family due to the containing relation that is. $\neg \underset{i,j;i\neq j}{\forall} (A_j \subset A_i) \vee (A_i \subset A_j)$

2. Sets that were defined in this way $\{A_i\}_{i=1\dots10}$ were made subject to selected transformations i_p so that for each i $B_{i,p}=i_p(A_i)$ which created a data base for the particular stages of the experiment (sets $B_{i,p}$ also do not form a declining family).

For the sake of the paper the domain of each classifier (and the mapping range i_p) will be called the input space respectively of: parameters, samples and coefficients

I. Tests of classifiers for parameters' space (A,E1,E2,DFR,V)

In this stage of the experiment mapping i_p (p=1) was given in the following way:

$$f \in C[a,b],\ i_1(f)=(A,E1,E2,DFR,V) \in R^5$$

where :

$A = \max_{x \in [a,b]} (f(x))$ - amplitude of envelope f

$E1 = \int_a^{t_1} f(x)^2 dx$ bottom hardness coefficient [1]

$E2 = \int_{t_1}^{b} f(x)^2 dx$ bottom roughness coefficient [1],

where t_1 is a point such that: $f(t_1) = A$

$DFR = \lim_{\Delta s \to 0} - \frac{\ln(N_f(\Delta s))}{\ln(\Delta s)}$ - fractal dimension of the envelope where N(Δs) - power of covering the envelope with squares at a side of Δs [3].

$V = \frac{V_a^b(f)}{(b-a)}$; - relation between envelope saltus and range length [a,b].

where:

$$V_a^b(f) = \lim_{\sigma \to 0} \sum_{i=1..n} |f(x_i) - f(x_{i-1})|$$

σ – radius of division: $a = x_1 \leq x_1 \leq x_1 \leq ... \leq x_{n-1} \leq x_n = b$ section [a,b]

In this given parameter space, a partial division of class D of echo envelope was observed (see Fig. 1), which led to the conclusion that the decisions made by the minimum-distance classifier should be sufficient in the task of bottom type recognition.

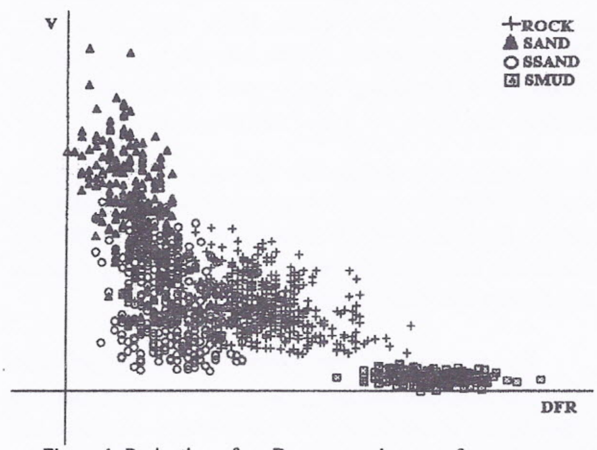

Figure 1. Projection of set $B_{6,1}$ over subspace of parameters (DFR,V)

Results of simulation for i_I, have been presented in the form of diagrams in Fig. 2 i Fig. 3. The visible "oscillations" of the diagram (especially in the case of the minimum-distance classifier) were caused by the fact that the training sets $\{B_{i,p}\}_{i=1...10}$ were not a declining family. However, the size of the oscillations provides additional information about how the quality of classification is dependent on the selection of the learning data. To estimate it what was calculated was the saltus V_{10}^{100} of diagrams of the obtained diagrams:

	min-dist	neural net	NEFCLASS
V_{10}^{100}	22.3	14.35	52.3

Table 1. Saltus of effectiveness diagrams of the analyzed classifiers for i_I

To be less precise, the saltus represents the number and size of diagram "oscillations", the higher the saltus, the higher the "oscillations" are. The results from Table 1 undoubtedly speak for the benefit of the neural network.

It is also worth noticing that the slight advantage of the minimum-distance classifier for big powers of the learning sets (see Fig. 2) can be misleading. It is known for a fact that the classifier on a training set always reaches 100% of correct decisions while it is difficult to perform approximation (as is the case with NEFFCLASS and the neural network) with an error small enough to come up with a result just as good.

In the course of learning the neural network the topology of the network was observed to have had a slight effect on the generalization effectiveness. Fig. 2 presents the results for the architecture with two hidden layers that contain respectively 5 and 4 neurons.

The test for the NEFCLASS classifier was not very successful. That was caused by the fact that for bigger training sets, the training of a classifier failed even though a big number of possible topologies were tested. A satisfactorily small error was obtained only in the case of sets $B_{10,1}$ i $B_{9,1}$ and in these cases (see Fig. 3 beginning of diagram) the quality of the classification is good.

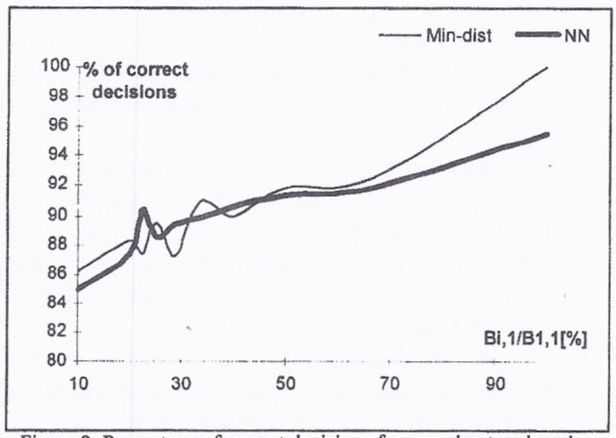

Figure 2. Percentage of correct decisions for neural network and the minimum-distance classifier depending on the size of the learning set for test I.

In general the results of the simulation confirmed earlier projections. The examined classes of sea bottom in the tested parameter space using Euclidean distance are well separable through the minimum-distance classifier. Basically it is sufficient for the purpose of this task. Even though the neural network is slightly less effective, this classifier seems to be less dependent on the choice of training data which makes it more useful in the classification systems of bottom sediments. The NEFCLASS classifier is clearly unable to handle the increasing training in spite of the promising tests for sets $B_{0,1}$ i $B_{9,1}$.

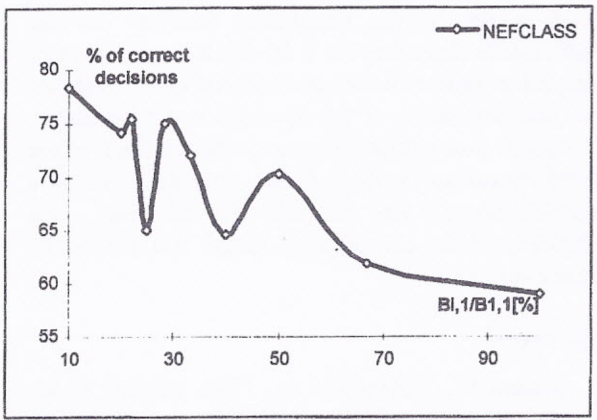

Figure 3. Percentage of correct decisions for the NEFCLASS classifier depending on the size of the learning set, for test I.

II. Tests for classifiers for a space of 50 samples of the envelopes

Mapping i_p (p=2) has been defined as follows:

$$f \in C[a,b],\ i_2(f)=(x_1,\ldots,x_{50}) \in R^{50}$$

where:

$$\underset{i}{\forall}\, x_{i+1} = f\left(a + i\frac{b-a}{49}\right), \qquad i = 0\ldots49.$$

The purpose of the test was to check the behavior of classifiers for a case when the input space is of a large size and the signal envelope was only sampled.

For a space of samples defined in this way, high effectiveness of the minimum-distance classifier as well as of the neural network classifier was observed (see Fig. 4). What might seem surprising is the high performance of the first one. What it means, however, is that the classes of bottom sediments can be easily separated in the space in question. In the case of the NEFCLASS classifier the learning process did not succeed (see Fig. 5) for none of the training sets. It undermines the usefulness of this classifier in those cases where its domain is a high dimension space.

By analogy to test I the saltus of the diagrams of the obtained relations was calculated:

	min-dist	neural net	NEFCLASS
V_{10}^{100}	15.3	13.2	35.0

Table 2. Saltus of effectiveness diagrams of the analyzed classifiers for i_2

Also in this case the neural network turned out to be a classifier that is less dependent on the selection of training data when compared to the minimum-distance classifier. Comments about the topology of the tested networks are similar to what they were in the first stage of the experiment. Fig. 4 illustrates the results for the architecture with three hidden layers containing respectively 20, 10 and 8 neurons.

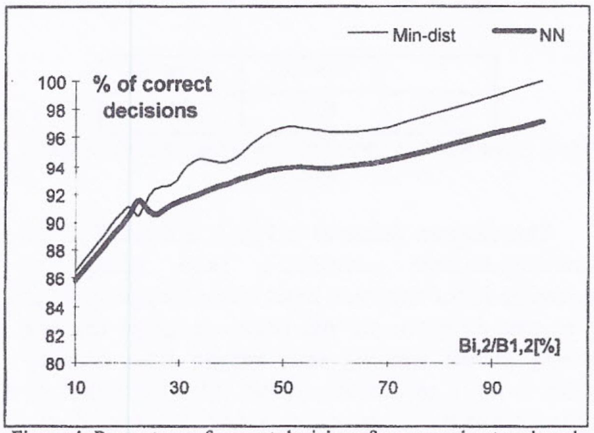

Figure 4. Percentage of correct decisions for a neural network and the minimum-distance classifier depending on the size of the learning set, for test II.

The general effectiveness of recognition of the first two classifiers is in this case higher than in test one. It is a direct consequence of a smaller reduction of information upon the usage of transformation i_2 instead of i_1.

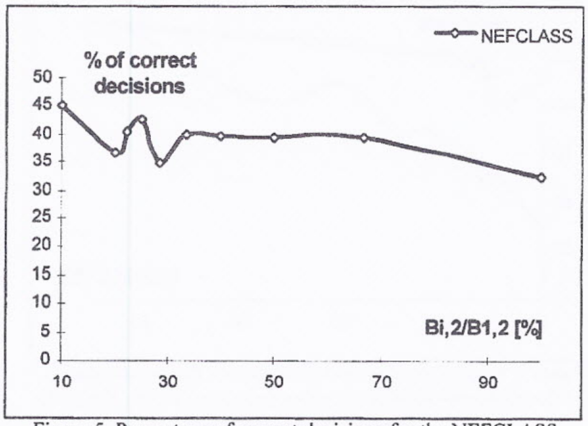

Figure 5. Percentage of correct decisions for the NEFCLASS classifier depending on the size of the learning set, for test II.

III. Tests for classifiers for a space of Fourier's coefficients

The purpose of the last test was to show that the minimum-distance classifier cannot always effectively separate the input space In this stage of the experiment the NEFCLASS classifier was not analyzed because the initial results turned out to be similar to those coming from previous tests..

This time i_p (p=3) was defined as:

$$f \in C[a,b],\ i_3(f)=(x_1,\ldots,x_{20}) \in R^{20}$$

where:

$$\forall_i x_i = \left| \frac{1}{b-a} \int_a^b f(t) e^{-j\omega(i-1)t} dt \right|, \qquad \omega = \frac{2\pi}{b-a}$$

	min-dist	neural net
V_{10}^{100}	63.6	24.3

Table 3. Saltus of effectiveness diagrams of the analyzed classifiers for i_3

The diagram obtained in Fig. 6 is a proof of the minimum-distance classifier's poor ability to generalize in the analyzed input space. The percentage of correct decisions for the entire set given the low powers of the training sets, ranges from 50-60% which is not a satisfactory result. Also the relatively high saltus of the diagram (see Table 3) is a proof of a higher dependency of the minimum-distance classifier on the learning data. Contrary to that the effectiveness of the neural network is comparable to the previous tests. The results illustrated in Fig. 6 were obtained for a two layer network of 8 and 10 neurons in the respective hidden layers.

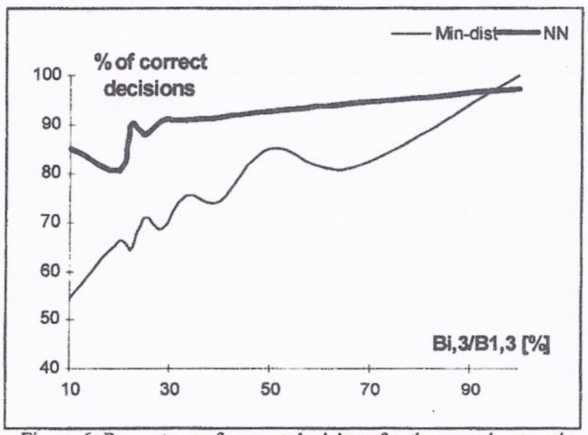

Figure 6. Percentage of correct decisions for the neural network and the minimum-distance classifier depending on the size of the learning set, for test II.

Conclusions

The experiment showed that the neural network is an effective classifier for the task of recognition the types of bottom sediments. This classifier also assures a satisfactory extrapolation of test sets.

Also the minimal-distance classifier was proved as an effective in the first two experiments and seems to be sufficient especially when the input space is a parameter space.

In the case of the NEFCLASS system, the only positive effect is the result for learning sets $B_{10,1}$ and $B_{9,1}$, in the other cases the process of training the classifier was not successful.

Good results of the simulation (leaving out the NEFCLASS classifier) for a 50 dimensional space of sampled envelopes could provide sufficient evidence that parametrization of the envelope is not necessary. This holds true especially because the topology of the tested neural networks is simple enough to allow a computation of the network in real time. The simplicity of the computation makes it a practically attractive tool.

References

1. Bakiera D., Stepnowski A., 1996, Method of the sea bottom classification with a division of the first echo signal, Proceedings of the XIIIth Symposium on Hydroacoustics, Gdynia-Jurata, 55-60.
2. Chivers R. C., 1994, Acoustical Sea-Bed Characterization, XIth Symposium on Hydroacoustics, Jurata.
3. Hastings H. M., Sugihara G., 1994, Fractals. A user's guide for the natural sciences, Oxford University Press, Oxford, New York, Tokyo.
4. Hertz J., Krogh K., Palmer R. G. 1995: Wstęp do teorii obliczeń neuronowych, WT, Warszawa
5. Koiran P., Sontag, E. D. 1994 Neural Networks with Quadratic VC Dimension.
6. Korbicz J., Obuchowicz A.,Uciński D. 1994, Sztuczne sieci neuronowe AOF, Poznań.
7. Nauck D., Klawonn F., 1996 Neuro-fuzzy classification initialized by fuzzy clustering, Proc. Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT96) Aachen.
8. Nauck D. and Kruse R. 1995 NEFCLASS - A neuro - fuzzy approach for the classification of data, Symposium on Applied Computing, Nashville.
9. Pouliquen E., Lurton X., 1992, Sea-bed identification using echo sounder signal, European Conference on Underwater Acoustics, Elsevier Applied Science, London and New York, 535.