# Independent Classifiers in Ontogenic Neural Networks for ATR

Janusz A. Starzyk
Ohio University

Dale E. Nelson
Wright Laboratory

**Abstract:**

*This paper investigates independence of classifiers in neural networks for high range resolution radar target recognition. Independent classifiers are used to select distinguishing features for synthesis of ontogenic neural networks (networks that generate their own topology during training). A class of nonorthogonal classifiers is defined and their classification properties are investigated. Radial basis functions and wavelet transforms are used to preprocess the radar signal data. Data preprocessing is used to minimize the effect of noise, phase shift, and scale change of the radar signal. Simulation results based on synthetically generated aircraft radar images showed promise for automatic target recognition.*

## Introduction

Automatic target recognition requires an accurate recognition algorithm which can be implemented on real time hardware. The problem of pattern recognition can be efficiently solved on a dedicated neural network (NN), trained off line, using an extensive data base. Neural networks use many paradigms and data organizations which are potentially useful for target recognition.

Radial basis functions (RBF) were shown to provide an arbitrarily close approximation to any continuous function [Girosi]. In addition, [Chen] demonstrated that any continuous function can be used as an activation function in RBF neural networks (RBFN), as long as it is not an even polynomial. Recently [Zhao] used RBFN for automatic target recognition based on Ku-band radar range profiles of three military aircraft. He demonstrated that by using a Fourier transform and non-coherent amplitude averaging, stable and shift invariant features were obtained which significantly improved the classification quality. In addition, classification based on RBFN was more accurate than using minimum-error Bayesian classifiers.

In this paper, we use RBF to produce a minimum set of independent classifiers in the input transformation space. The input transformation space includes the original signal, its amplitude range, average value, standard deviation, and additive information cost functions (Shannon entropy, log energy, and $l^p$ norms.). In addition a Fourier transform and a Haar wavelet transform of the original signal were used to enhance the input transformation space. In this space, a selection of features for nonorthogonal classifiers were performed. The classification results were compared with classification based on the orthogonal RBF classifiers obtained in the same input transformation space. NN training and test were performed using model based, synthetically generated, high range resolution radar aircraft images.

**Classifiers**

A fundamental problem in pattern classification is to determine a class membership with the maximum statistical confidence of the correct classification decision. While this classification task can be performed with 100% probability for the trained data, there is no proof that a neural network capable of 100% recognition of the trained data yields a similarly high recognition rate for test data which were not used during training. To the contrary, such networks are quite often unable to classify new test data and are known as overtrained networks unable to generalize. In addition, the overtrained neural networks are usually designed without regard to the hardware cost and result in an excessive architecture for a given task. Therefore, when selecting a neural network for pattern classification and its training algorithm, one must consider tradeoffs between the need to separate members of different classes and the need to unite members of the same class, as those needs translate into contradicting hardware requirements. Unification of a large number of training data into a single decision making operation usually translates into simplification of the classification task and produces a NN capable of a useful generalization, while separation makes it more complex and reduces its generalization ability.

In order to facilitate the classification task, distinguishing target features are used. These features are selected from a predefined set of the transformation functions of the input data. Depending on the task and imagery data, the separation boundaries between different classes will be crisp or fuzzy, leading to different types of neural network processing components and different types of membership functions of the resulting classification. To construct a feature selection mechanism, we define a measure of performance relative to the classification task. We will represent a given pattern as an ordered set of values. For instance, a binary image can be represented by a vector of 0 and 1 values arranged in the specific scanning order or a Fourier spectrum of a radar signal can be represented by a vector of complex values at discrete frequency points.

To compare and classify patterns using such a vector representation and to preserve all the useful information before a feature selection criteria are established, we assume that in a given classification task all vectors have the same dimensions. If needed, vector sizes may be extended to match the longest vector in the input pattern set. Vector padding, sampling, time, and frequency scaling may be used to satisfy this requirement. A linear vector space which contain all the vectors obtained from the input pattern set is called the ***input space***. Input space may either contain raw input data or some functions of these data. By allowing initial transformation of the input data we have greater flexibility to formulate input space. A direct representation of the input signals and its transforms yields vectors of high dimensionality. To facilitate the classification task, the transformation of the input space to a space of a lower dimension is used and is optimized for a specific classification task.

A specific transformation of the input is referred to as the ***input feature***. Using an unlimited number of features one can achieve a linear separability of any input data, which is critical for any classification problem. But, this selection may lead to a costly classification procedure and hardware. In addition, it will not be useful for classification of new data. By proper selection of the input features we can obtain better classifiers, resistant to noise and local image distortions, and applicable to different types of sensor data.

A mathematical task of checking the independence of functions or vectors is easily accomplished by using linear algebra methods. On the other hand, there is no mathematically

formulated method to check the independence of features selected for the classification task. This makes design of optimum software and hardware methods for pattern recognition hard, and also makes any comparison between different classifiers difficult. In this research, we will define a way to check the independence of features used for pattern classification. By using independent features, a systematic way of synthesizing minimum classifiers will be derived, which will permit the design of optimized architectures of the resulting NN hardware.

Since the individual features will be defined to optimize the classification task, the obtained classifiers will be optimum in the sense of the selected transformations, so they are expected to outperform any ad-hoc classification technique. The resulting optimized classifiers will perform their classification task with a high tolerance to deviations within a specific class. They will be able to avoid incorrect classifications and minimize the false alarms better than ad-hoc designs.

**Feature selection for NN pattern recognition**

To learn a nonlinear mapping from the input space to the output space, one needs to consider independent transformations of the input space. Such transformations can be easily obtained using either a complete set of orthogonal functions, in which orthogonality guarantees transformation independence, or using a successive approximation of the learned mapping, in which each successive transformation is found by orthogonalizing the error of the existing fit. Similar to this operation is orthogonalization of vectors in a linear vector space. By using Gram-Schmidt orthogonalization we successively remove the components of the input vectors which are linearly dependent on the selected orthogonal set of vectors. In order to obtain an efficient classifier, a similar operation needs to be defined on a set of input transformations (features). This operation would tell us how independent a classification based on a selected feature is as compared to the previously selected features. By introducing this operation and a measure of separation ability we will define a tool for selection of the dominant features for building the optimum classifiers.

To this end let us define a ***feature*** *f* mathematically as an ordered pair $(F, \Omega)$ of a nonlinear transformation *F* and a proper subset of its output space. We define a ***feature domain*** *D* as a subset of the domain of the transformation *F* which is mapped into $\Omega$, and a ***feature sample set*** *S* as a subset of the input training data included in *D*. For instance a cluster of points in the input space is defined by a feature in which a nonlinear transformation *F* is the Euclidean distance from the cluster center and the proper subset $\Omega$ is the closed interval [0, R], where R is the cluster radius. The feature domain *D* is a sphere with radius *R* located at the cluster center, and the feature sample set *S* is equal to the set of input points included in *D*.

Notice, that for a given transformation *F* we can define infinitely many features by simply modifying the subset $\Omega$. An entire classification task can be based on a single transformation *F* paired with different sets $\Omega$. For example, a clustering method can be used to classify a number of patterns based on the nearest neighbor rule and a number of clusters defined in the input space.

Definition: A feature $f_m$ is ***covered*** by the features $f_1, f_2, \ldots, f_k$ iff

$$D_m \subset D_1 \cup D_2 \cup \ldots \cup D_k .$$

Definition: A set of features $\Phi = \{ f_1, f_2, \_, f_n \}$ is ***independent*** if none of its elements can be covered by others.

Consider a set of *training data T* used for the task of pattern recognition. This set is composed of subsets of vectors from different classes $T = C_1 \cup C_2 \cup ... \cup C_c$, where $C_i$ is a set of vectors from the class *I*. For simplicity we will use the same symbol to represent a class and its set of input vectors. Without loss of generality let us assume that all the classes are disjoint, i.e. $C_i \cap C_j = \varnothing$ for $i \neq j$.

Definition: A feature $f(C_i)$ is a ***differentiating feature*** of class $C_i$ if its domain includes only the input vectors from the class $C_i$.

Notice, that whether a feature is a differentiating one or not depends on the complete training set *T*. Adding new training data may change a differentiating feature to become nondifferentiating and removal of a training data may change a nondifferentiating feature to differentiating one.

Definition: A set of differentiating features $\Phi(C_i) = [\, f_1(C_i), f_2(C_i), ..., f_n(C_i) \,]$ is an ***orthogonal classifier*** for class $C_i$ if the sum of domains of its features includes the set $C_i$. The classifier is a ***minimal classifier*** if the differentiating features are independent.

In general, to classify input data we can use a set of orthogonal classifiers defined for all classes. Using orthogonal classifiers will yield 100% recognition for the training data. In addition, orthogonal classifiers will result in a simple two level structure of the classification neural network. However, the recognition rate for new data may be significantly smaller, as quite often the orthogonal classifiers are based on a large number of independent features, which leads to small feature domains and poor generalization ability. In addition, as a result of a large number of features used, the hardware requirements for the neural network based on orthogonal classifiers are high. The network will use features defined in the input space or the transformed input space, and it will have a two level structure with a high demand on wiring the input data.

In order to design a simple classifier we must find a set of independent features of minimum cardinality which differentiate all classes. In this paper we introduce feature selection based on a ***sequential classifier***. A sequential classifier is obtained as follows: A differentiating feature is selected and its sample set is removed from the input space. Then, another differentiating feature is selected and its sample set removed. This process is repeated until all training samples are classified. A sequential classifier is an example of a nonorthogonal classifier. In this work we demonstrate that nonorthogonal classifiers are able to obtain correct classification of the trained data with better generalization ability than orthogonal classifiers. Sequential classifiers are just one example of nonorthogonal classifiers. They will result in a multilevel NN structure in which both number of neurons, processing layers and the overall organization is a function of the input data, which is a characteristic feature of the ontogenic NN [Ensley].

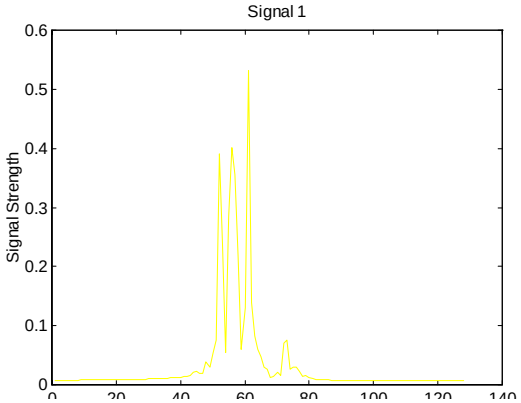
Since an input vector may or may not exhibit an individual feature, we can design combinational classifiers in which a decision regarding an input classification is expressed by a combinational logic function which depends on several features. This leads to the construction of pattern recognition neural networks in which classification decisions are made by a network of logic gates. Such classifiers will be extremely hardware efficient, as a multilevel logic synthesis could be used to synthesize its structure for optimum cost and performance. Combinational classifiers will be designed using complex logic structures of the decision making

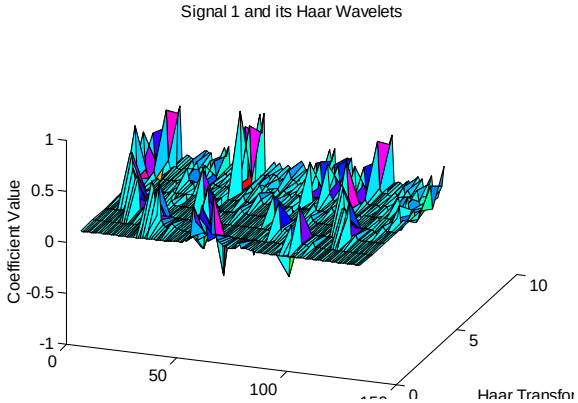
process. This will yield a multilevel NN structure. Probabilistic or fuzzy classifiers can also be designed by using fuzzy logic instead of binary logic.

**Simulation results**

In this work, sequential classifiers are used to demonstrate a potential utilization of nonorthogonal classifiers for automatic target recognition. A simple feature selection method was used to demonstrate that the nonorthogonal classifiers work even with the simplest thresholding features. It is expected that, when used with a more elaborate feature selection process, nonorthogonal classifiers will maintain their advantage over the orthogonal classifiers. The experiments described in this paper used synthetically generated high range resolution radar (HRR) data. The full data set has six air-to-air targets. Each target set consists of 1071 range profiles with each range profile consisting of 128 range bins. The target pose is head-on with a range of azimuth of ±25° and elevations of -20° to 0°. From the complete data set, a training and test set were extracted. Each of the training and test sets consisted of 60 randomly selected range profiles for each target. For each target in these sets the same azimuth and elevation range profiles were used. This yielded 360 range profiles for each data set. The range profiles were selected such that the training and test sets were disjoint. The reason that these data sets are so small is to speed the computations for a proof of concept. In actual usage, much larger training and test sets would be used. It should also be noted that it is very possible that the test data ranges outside the training data set. Thus causing the methodology to extrapolate and may result in lower recognition rates.

The input space includes the original signal, its amplitude range, average value, standard deviation, and additive information cost functions (Shannon entropy, log energy, and $l^p$ norms.). In addition a Fourier transform and a Haar wavelet transform of the original signal were used to enhance the input space. In this space a selection of features for orthogonal and nonorthogonal classifiers were performed. Fig. 1(a) shows an example of a raw signal data, and Fig. 1(b) its wavelet transform.




(a)A sample HRR signal Figure 1 (b) Haar wavelet of the selected signal.

Figure 1

Feature selection was based on the *slicing approach* in which input data were projected onto one dimensional subspaces. In these subspaces, intervals which include input vectors from different classes were found. These intervals define slices in the multidimensional input space

which contain only the samples of a single class. Features were selected based on the cardinality of input vectors included in these slices. Fig. 2 illustrates maximum cluster sizes for different transformation functions of the input data. We may observe that the best features are based on the signal transforms rather than on the raw data (represented by features 7-134.) This is in agreement with other research results in ATR, which indicate that preprocessing may enhance classifiers' recognition ability. Other feature selection methods can be applied at this stage. For instance the mutual information approach with its efficient search algorithm [Battiti] can be a useful alternative. It will be used in formulation of the combinational classifiers.

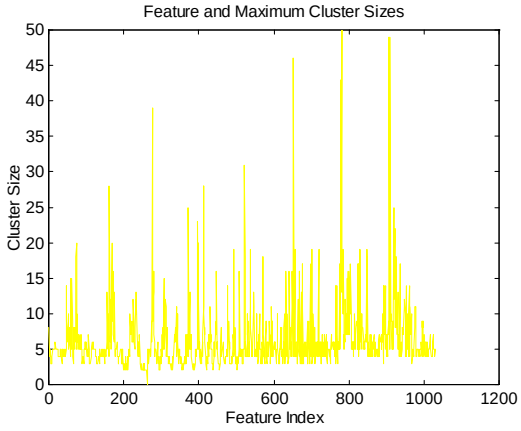



Fig.2 Different features and their maximum cluster sizes.

An orthogonal classifier was obtained by using an equal number of differentiating features for each class. Differentiating features were selected based on the largest sample sets. A sequential classifier was used as an example of nonorthogonal classifier. Differentiating features were also based on the largest sample sets. Since sequential classification procedure changes the input space, after each step the projection intervals and cardinality of input vectors in various slices changed, effecting the choice of differentiating features. In general, the obtained sequence of features may have a nonmonotonic cardinality. This dynamically changing cardinality of input features is illustrated in Fig. 3.

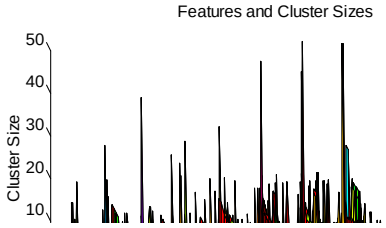



Fig.3 Features' cardinality in the selection process.

After feature selection was completed a RBFN was obtained for each classifier. This results in three layer NN for both orthogonal and nonorthogonal classifier. This was done to demonstrate differences in the classification performance of the two classifiers which result from differences in the feature selection process, rather than NN organization. Normally, nonorthogonal classifiers will be realized in a multilayer NN which may lead to simpler hardware. The results of classification are shown in Table 1.

| | Training Data | | | | Test Data | | | |
|---|---|---|---|---|---|---|---|---|
| | Orthogonal | | Sequential | | Orthogonal | | Sequential | |
| Target Number | Recognition Rate % | Error Rate % | Recognition Rate % | Error Rate % | Recognition Rate % | Error Rate % | Recognition Rate % | Error Rate % |
| 1 | 87 | 0 | 83 | 0 | 83 | 0 | 85 | 0 |
| 2 | 53 | 0 | 60 | 3 | 28 | 18 | 32 | 8 |
| 3 | 50 | 0 | 63 | 0 | 40 | 2 | 73 | 3 |
| 4 | 37 | 0 | 67 | 3 | 25 | 18 | 53 | 15 |
| 5 | 28 | 0 | 78 | 2 | 12 | 0 | 48 | 10 |
| 6 | 28 | 0 | 48 | 2 | 18 | 7 | 35 | 15 |

Notice, that the classification results were obtained with a small number of simple feature functions. The sequential classifier used 29 features and the orthogonal classifier used 30 features to classify targets into 6 different classes. Typically, a much larger selection of features or much more complex features are used. In addition, the training set was deliberately small to provide a proof of concept for the enhanced performance of nonorthogonal classifiers.

**Conclusion**

This paper investigates independence of classifiers in neural networks for high resolution radar target recognition. Sequential classifiers were used as an example of nonorthogonal classifiers to select distinguishing features for synthesis of ontogenic neural networks. An orthogonal classifier based on the dominant distinguishing features was selected to compare the classification performance. Wavelet transforms and other signal transformations were used to preprocess the radar signal data. Simulation results demonstrate a potential benefit of using nonorthogonal classifiers for automatic target recognition

**References**


[1] F. Girosi and T. Poggio, “Networks and the best approximation property,” Artificial Intelligence Lab, Memo 1164, MIT (1989).

[2] T. Chen and H. Chen, “Approximation capability to functions of several variables, nonlinear functionals and operators by radial basis function neural networks,” IEEE Trans. Neural Networks, vol. 6, pp. 904-910, 1996.

[3] Q. Zhao and Z. Bao, “Radar target recognition using a radial basis function neural network,” Neural Networks, vol. 9 no. 4, pp. 709-720, 1996.

[4] D. Ensley and D. Nelson, "Applying cascade correlation to the extrapolation of chaotic time series," Proceedings of the Third Workshop on Neural Networks: Academic, Industrial, NASA, Defense 92; (Auburn AL, Feb. 1992).

[5] R. Battiti, “Using mutual information for selecting features in supervised neural net learning”, IEEE Trans. Neural Networks, vol. 5, pp. 537-550, July 1996.