Theoretical and Computational Acoustics
D. Lee, Y-H Pao, M.H. Schultz, and Y-C Teng (Editors)

On the Detection and Identification of Marine Mammals Using Artificial Neural Networks

Christopher M. DeAngelis, Kurt R. Almquist and Kelly J. Ross
NUWC Division Newport, Code 2253
Newport, Rhode Island 02841

Abstract: A technique for detecting and identifying marine mammals is presented. This technique utilizes a wavelet based preprocessor to extract features from a 1/3 of a second marine mammal sound clip. The dominant features are then used by an artificial neural network to identify the species of the marine mammal. In addition, a technique for unwrapping the internal mappings of the artificial neural network is presented. A combination of backpropagation, neural net weight decay and pruning are used to assist in forming a minimum neural network architecture. This resulting architecture can then be decomposed and analyzed, determining internal mappings, hidden node functionality, as well as input node importance and utilization. Furthermore, the actual features needed to determine and discriminate the classes can be identified as a function of classification accuracy. Results show that modest classification can be achieved with only a minimum number of features, and a higher degree of classification requires the non-linear combination of more features. The key is selecting the correct ones. The presented technique provides this ability.

1. INTRODUCTION

This research developed a technique for distinguishing between species of marine mammals. Initially, the technique uses a wavelet-based preprocessor for feature extraction, followed by an Artificial Neural Network (ANN) to perform the differentiation between species of marine mammals. It is planned to expand the ANN to recognize more classes of marine mammals in the future.

The overall goal of this research is to develop a technique for unwrapping and understanding the internal mappings of a backpropagation ANN. This will make it possible to validate the operations of the network, and determine the bounding envelopes under which it will correctly operate. By determining how the inputs are being combined and processed when inputs are partly obscured or completely degraded, it is possible to attain insight into the ANN's performance under severe operating conditions.

1.1 Bounds of Research

This research focuses on classification applications with boolean outputs in order to simplify the network decomposition process. For example, it is easier to determine which nodes contribute to a boolean output node turning *on* or *off*, than it is to determine the cause of a 5% increase in a continuous valued output.

To further simply the network decomposition process, the research began by examining only single output architectures, thus all of the hidden nodes contribute to the value of a single output. If multiple outputs were processed by a single ANN, then each node could influence multiple outputs, and a scheme would have to be developed to separate and decouple the contribution of each node to each of the output nodes. Once a technique is developed for understanding mappings for the single boolean output case, it can be expanded to accommodate multiple, as well as continuous valued, outputs.

1.2 Technique Overview

To facilitate the generation of minimum sized neural networks, a number of algorithms were exploited. First, wavelets[1] are used to preprocess the sounds. This allows for accurate representation of the signals with a minimum number of terms. These terms are used as inputs to the neural based classifier. Next, a weight decay algorithm[2] was incorporated with a quasi-Newton weight update algorithm in the backpropagation ANN, encouraging unnecessary terms to fade to zero and enhancing error surface minimization. Finally, a network pruning algorithm[3] is used to remove nodes and interconnections that are not required for the mapping, with any minimal or redundant contributions redistributed to other active nodes. Collectively, these algorithms promote the reduction of a signal to a representative set of features that can then be combined by a minimally sized ANN to form the class mappings. The process of decomposing and interpreting the underlying mapping is greatly simplified by the use of a minimal-sized ANN architecture.

Since the sounds being classified contain little periodic information, a wavelet preprocessor was selected instead of the Fast Fourier transform (FFT) or the windowed FFT. Unlike the FFTs, wavelets more naturally capture the time-varying characteristics of transient signals, allowing the signal's energy to be accurately represented with a minimum number of coefficients[4,5].

Neural network based data processing systems have a number of favorable characteristics[6,7]. In contrast with many classical systems, they are adaptable, fast in execution, require no a priori statistics, can accommodate non-stationary data, and have the ability to approximate system functions that are not well definable or are even unknown[8,9]. These characteristics make them well suited for pattern completion and pattern matching applications[10,11].

If an Artificial Neural Network (ANN) can be reduced to a minimum architecture, then the weights of the resulting sparsely connected nodes can be more easily examined to determine node functionality. In this paper, a minimum architecture is defined as one providing acceptable performance with a minimum number of nodes and interconnections. Node functionality refers to the contribution each node is making to the overall mapping. By analyzing the connections from the input layer to the hidden layer, the importance and use of each input can be determined. The importance of an input node is reflected in the number of hidden nodes actually using that input. Determining the usage of an input is a bit more tenuous, as this information must be extracted from the weights assigned to each connection between the input and hidden nodes.

The weights of a network contain the information on how the inputs are combined to generate the desired output. An advantage of having a minimum architecture network is that all of the nodes are being used to their full capacity. One attribute of full capacity refers to a node having a dynamic output value which varies as the node's inputs change. Full capacity also means that the nodes have their own identity and are not mimicking or duplicating other nodes. When each node is used to its full capacity, one can determine whether it is adding to or detracting from the evidence that class "A" or class "B" exists by looking at the polarity of a hidden node's connections along with that of its bias term.

2. ANALYSIS

2.1 The Data Set

The data set being used consists of multiple samples of underwater acoustic signals. These signals have been sampled at 25KHz and each contain nearly a million data points. The four classes contained in these signals are: Carpenter Fish (also known as a Sperm Whale), Dolphin, Porpoise, and an underwater vehicle. It should be noted that in the case of the marine biological signals, a number of different animals were recorded for each species. It is hoped that this will promote the classification of the species as a whole and not just the individual members in each class. Additional data is needed to validate this premise since there is no indication in the data at hand of when each individual is speaking. Numerous individuals were present during each recording.

800 sound clips of each class were extracted from the recorded samples using a 8192 point window, resulting in approximately one-third of a second of information. Half of these clips were used for training the neural network and the other half for testing. The clips were not extracted in any particular order in that a random window placement was performed.

2.2 Preprocessing The Data Using A Wavelet Transform

Each of the sound clips was preprocessed using a Daubechies 4 wavelet. Due to the nature of the wavelet transform, the number of wavelet translations and dilations is equal to the length of the vector being analyzed which was 2^{13}. Therefore, the 8192 point vector is translated into 8192 wavelet coefficients, the majority of which are zero. These coefficients are then normalized such that the largest is 1. When the coefficients were sorted from largest to smallest, a rapid exponential decay in the coefficient magnitudes was noticed, meaning that most of the signal's energy could be represented by a few coefficients. In order to retain the time and frequency information associated with each coefficient, an additional vector is created containing the respective wavelet matrix entry locations. The largest 20 coefficients along with their 20 respective time and frequency locating indexes (40 terms in all) are retained and used as inputs to the neural network.

2.3 Classification Of The Data With An Artificial Neural Network

In the first stage of this research two species were chosen to classify: porpoise and carpenter fish. The respective assigned class values are 0 (zero) and 1 (one), providing the single output architecture needed to simplify node analysis. The follow-on to this work will expand analysis to the multiple output case where each output will represent an individual class.

The backpropagation supervised learning neural paradigm[12] was used. This network "learns by doing" which means when the ANN is repeatedly presented with input-output training vectors, its internal connections adapt according to a given cost function until the correct response is given for a respective stimuli[13,14]. When analyzing the performance of this network, a classification threshold of 0.5 is used. Network output values greater than 0.5 are said to be in the 1 class and the others in the 0 class.

Since the sizing of the network is ad hoc, a single hidden layer of 10 nodes was chosen, resulting in a 40 input, 10 hidden, 1 output neural net architecture (40x10x1). This gave 421 degrees of freedom (the number of adjustable weights and biases) to separate the two classes. A number of networks were trained, each starting with a different set of random weights. All converged and correctly classified over 95% of the sounds for each class.

2.4 Neural Network Minimization

To aid in the ANN unwrapping process, several techniques were used to reduce the size of the ANN from the original 40x10x1 size. When analyzing the activation levels of the trained hidden nodes to various inputs, a great deal of duplication was discovered. The levels indicated that only 4 of the hidden nodes were actively contributing to the mapping. Thus, new ANNs with a 40x5x1 structure were then created. These networks were trained and gave similar, but inferior, classification results. Although there were still enough degrees of freedom, it was not surprising that the 40x5x1 networks did not perform as well as the 40x10x1 networks since fewer degrees of freedom means less of a chance to converge to a good solution.

An automated method to reduce the network's architecture was needed. The implemented minimization technique used a weight decay algorithm[2] incorporated into the quasi-Newton weight update scheme followed by an Optimum Brain Surgeon[15] based pruning algorithm[3]. The pruning algorithm was utilized to remove the non-needed nodes having constant activation levels, as well as nodes contributing little to the overall mapping. The contributions of any deleted nodes were redistributed over the remaining nodes. In addition to the nodes removed by the pruning algorithm, any of the remaining interconnections between nodes that do not contribute to the mapping are also removed. Some additional training is required after connections are pruned in order to refine the mapping.

Test set classification increased to 97% when this iterative process of decaying, pruning, and training was employed on a 40x10x1 network. When analyzing the remaining 102 interconnections, it was noticed that a considerable number of connections to the inputs corresponding to the wavelet coefficients were removed, leaving the majority of the connections to the time-frequency input nodes. This suggests that the time at which various frequencies are present is more important in separating the two classes than the magnitude of the frequencies. It was also noticed that the separation between the classes was within 0.1 of the extremes far from the 0.5 threshold value (i.e. the output of the network was less than 0.1 for all of class "0" and greater than 0.9 for all of class "1". This indicates that there is a strong separation between the species of marine mammals classified by the neural network.

2.5 Unwrapping The ANN Internal Mappings

Once the minimally sized network that provides the desired class separation was found, we can proceed to determine which inputs are important for classification, and how they are combined to provide a mapping for each class. The size and architecture of the network suggests the complexity of classification, with the number of connections to each input reflecting that input's importance and the polarity of the interconnections encouraging or inhibiting a hidden node's activation.

In order to determine which inputs are fundamental to class separation, one of the correct classification 100% networks (on the training set) was pruned until 50% correct classification was no longer achievable. Pruning was performed in stages with the number and location of pruned interconnections saved along the way. The smallest network that provided better than 50% classification contained 6 parameters (from 102), connecting 2 hidden nodes to 3 inputs. The resulting classification was better than 72% combining only 1 coefficient with 2 time-frequency indexes. More analysis needs to be conducted, but it appears that the output node is biased *on* with the two hidden nodes providing evidence to turn it *off*. Furthermore, the inputs need to be un-normalized and analyzed to determine the respective time-frequency magnitude and locations. This may suggest that in this two class application, a simple band-pass filter could be used to separate the classes; looking for energy in a specific frequency band.

Remember that more interconnections are needed to improve this classifiers performance, but only 72 of the 102 are needed for 92% classification. As connections are added back to the network, their polarity and what connecting nodes give insight into the rebuilding of the network's internal mapping. The functionality of the individual nodes should not change, only the accuracy in the mapping.

3. FUTURE WORK

Additional network analysis needs to be conducted in order to determine the behavior of the network (i.e. What each node is doing and how it contributes to the overall network mapping) which will be done by examining node activation levels for each class. This will allow the determination of which inputs are used for classifying each class, and which inputs are taken collectively to represent a class or discriminate between classes. Once this is done, these results will be used to determine what functionality the nodes are providing in terms of known functions, such as band-pass filters or differentiators.

As parameters are added back into the network we can continue to refine our notion of how inputs are being put together to represent or separate the classes: Which nodes are acting as band-pass, low-pass, high-pass, or feature-pass filters? Are there feature or class differentiating nodes? These questions can be answered as the network is built back up to the acceptable classification level.

4. CONCLUSIONS

A technique for identifying several classes of marine mammals has been developed along with a technique for unwrapping the internal mappings of a backpropagation neural network for classification applications. This involves the use of a wavelet transform to extract the most representative features from various classes of signals. The largest 20 wavelet coefficients and their respective time-frequency indexes are used as the inputs to a neural network classifier trained to separate the signals of interest.

An iterative process of training the neural network, decaying the weights, and pruning the non-important nodes and interconnections from the neural network was employed to form the smallest possible network providing the desired level of classification. In all cases, it was apparent that the wavelet indexes are more important in providing class separation than the wavelet coefficients themselves. In addition, as the network was pruned further it was determined that the output node was initially biased toward one class and the hidden nodes provided evidence suggesting the other.

Additional research needs to be conducted to compare the actions of hidden nodes to more understandable functions, such as band-pass filters. In addition, once the boolean output case is understood, it is hoped that this technique can be extended to multiple, continuous-valued outputs.

Acknowledgment: This research was sponsored by ONR (Dr. J. Davis) and is part of an ongoing effort at NUWC Division Newport, Newport, Rhode Island. The data and its wavelet transformation was provided by Quyen Huynh and Walter Greene, Weapons Technology and Undersea Systems Department at NUWC. The weight decay and pruning algorithms were provided by Robert Green, Computer and Information Science Department, University of Massachusetts Dartmouth. The principal investigator for the technique of decomposing and unwrapping the neural network is Christopher DeAngelis, Combat Control Systems Department at NUWC.

REFERENCES

[1] C.K. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, Ca, 1992.

[2] A. Krogh and J. Hertz, "A Simple Weight Decay Can Improve Generalization", Advances in Neural Information Processing 4, 950-957, Morgan-Kaufmann, 1992.

[3] D. Weigand, D. Rumelhart and B. Huberman, "Generalization by Weight-elimination with Application to Forecasting", Advances in Neural Information Processing 3, 875-882, Morgan-Kaufmann, 1991.

[4] B. Jawerth and W. Sweldens, "An Overview of Wavelet Based Multiresolution Analysis", SIAM Review 36, No. 3, 377-412, September, 1994.

[5] J. Benedetto and M. Frazier, Eds., *Wavelets: Mathematics and Applications*, CRC Press, Boca Raton, 1993.

[6] M. Minsky and S. Papert, "Perceptrons: An introduction to Computational Geometry," MIT Press, Cambridge, Massachusetts, 1969.

[7] R.P. Lippman, "An Introduction to Computing with Neural Networks," IEEE ASSP Magazine Vol. 4, No. 2, 4-22, April, 1987.

[8] W.T. Illingworth, "Beginners guide to neural networks," IEEE AES Magazine September, 44-49, 1989.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, Macmillan College Publishing Company, New York, 1994.

[10] R.P. Lippman, "Pattern Classification Using Neural Networks," IEEE Communications Magazine, November, 47-64, 1989.

[11] E. Wilson, "Transient Detection and Feature Extraction Using Neural Networks," University of Rhode Island PhD Thesis, 1993.

[12] J. L. McClelland and Rumelhart D. E., *Explorations in Parallel Distributed Processing,* MIT Press, Cambridge, Massachusetts, 1988.

[13] D.E. Rumelhart, J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing, Volume 1: Foundations* , MIT Press, Cambridge, Massachusetts, 1989.

[14] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Harvard University Ph.D. Thesis, dated 1974.

[15] B. Hassibi and D. Stork, "Second order derivatives for network pruning: Optimal Brain Surgeon", Advances in Neural Information Processing, Morgan-Kaufmann, 164-171, 1992.