AN INTEGRATED HYBRID NEURAL NETWORK AND HIDDEN MARKOV MODEL CLASSIFIER FOR SONAR SIGNAL CLASSIFICATION

Amlan Kundu and George C. Chen

Naval Command, Control and Ocean Surveillance Center
RDT&E Division
San Diego, CA 92152-5001 USA

ABSTRACT

In this paper, we have presented an integrated hybrid neural network and hidden Markov model (HMM) classifier that combines the time normalization property of the HMM classifier with the superior discriminative ability of the neural net (NN). Sonar signals, like speech, display a strong time varying characteristic. Although the neural net has been successful in classifying transient like sonar signals, the success is achieved either by using a bigger net architecture or by incorporating a detection mechanism in the classification procedure. In this paper, we have proposed an integrated hybrid HMM and neural net classifier where a left-to-right HMM module is used first. The HMM module segments the observation sequence belonging to every exemplar into a fixed number of states starting from the left. After this segmentation, all the frames belonging to the same state are replaced by one average frame. Thus, every exemplar, irrespective of its time scale variation, is transformed into a fixed number of frames, i.e., a static pattern. The multi-layer perceptron (MLP) neural net is then used as the classifier for these time normalized exemplars. For successful modeling and classification, each frame is succinctly represented by a feature vector. Two feature extraction schemes are considered in this paper -- the first one is based on FFT power spectral coefficient, and the second one is based on quadrature mirror filter (QMF) bank based subband decomposition. Finally, some experimental results are provided to demonstrate the superiority of the hybrid integrated classifier.

1. INTRODUCTION

Some recently completed works [1, 2] have demonstrated the success of hidden Markov model and other classifiers in classifying the oceanic transients. Moreover, the results reported in [2] has established that the recognition results would be better and more robust by combining the evidence of both HMM and NN classifier in the decision making stage. The present work intends to build the foundation of one unique classifier that would incorporate the theoretical and practical advantages of both HMM and NN classifier in the classifier itself, i.e., build ***one classifier*** that would handle wide temporal variability and would provide strong inter-class discriminative power. In [3], a hybrid HMM/NN classifier is described. The scheme in [3] uses the learning vector quantizer (LVQ) as the neural net. In our view, the multi-layer perceptron neural net (MLP-NN) provides more discriminative power vis-a-vis the LVQ as it uses a hidden layer of nodes to provide nonlinear hyper boundaries separating the classes in the decision space. We also have incorporated the modified Viterbi algorithm in the HMM scheme to provide full state segmentation, as described in the sequel, of the signals. Without such full state segmentation, the hybrid classifier will simply fail to operate.

2. FEATURE SELECTION

We have two different feature representation schemes: one based on Fourier power spectra, and the other based on the QMF subband based decomposition.

2.1. Feature Selection from QMF Subbands

After the signal is decomposed into L subbands, the root mean square energy in each subband is computed. The two lowest frequency subbands are not used because the signals in these bands are due to sonobuoy vibration noise. The *rms* energy in the bands are arranged into a feature vector to represent the signal.

2.2. Feature Selection from Fourier Power Spectra

From the given data segment, its FFT is computed. Before FFT computation, each data segment is windowed with a Kaiser--Bessel window function. The magnitude square of the FFT coefficients gives the Fourier power spectrum of the data. Since some frequencies may be more useful than others, only the important frequencies should be selected for a compact representation of the signal for classification purpose.

3. CLASSIFIER DESIGN

0-7803-2431-5/95 $4.00

In our work, we have combined two classifiers: HMM and MLP-NN.

3.1. Hidden Markov Model

For each class, one HMM is designed. So, if there are M classes, M HMMs are designed [2]. Assume that we have M models denoted by λ_m, m=1,2,...,M. Given an unknown exemplar O, we calculate the maximum likelihood function against all the models using the Viterbi algorithm, i.e., we first calculate $P(O,Q^* \mid \lambda_m)$ for λ_m, m=1,2,...,M. Here, Q^* stands for the optimal state sequence given by the Viterbi algorithm. The exemplar O is recognized as belonging to the best matched model. In our task, we need the optimal full state sequence rather than the optimal class. Fortunately, the modified Viterbi algorithm provides the optimal full state sequence.

3.2. Multi Layer Perceptrons

Multi-layer perceptrons (MLP) are feed-forward nets with one or more layers of nodes between the input and output layers. A three-layer perceptron is used in our scheme. Generally, the multi-layer perceptrons are trained with the error back-propagation (EBP) algorithm which is an iterative gradient algorithm designed to minimize the mean square error (MSE) between the desired output y_k^* and the actual output y_k. A momentum term is also included in the training procedure. The details of this algorithm can be found in [4].

3.3. Integrated Hybrid Neural Net and HMM Classifier

Our integrated classification scheme is shown in Fig. 1. For each signal class, one L-R HMM is designed. During training of the neural net, each exemplar is first scored against all the models. The model which provides *the highest score as well as full state segmentation* is selected. By full state segmentation, we mean that the observation sequence, after being labeled by the Viterbi algorithm, has all the states present. Once a sequence has full state segmentation, all the frames in that sequence belonging to the same state is replaced by an average frame. Thus, if each HMM is designed with N states, after this frame normalization scheme, each exemplar is replaced by N frames irrespective of its time scale. If each frame is represented by an M dimensional feature vector, NM features are used as input to train the MLP neural net.

During classification, each exemplar is similarly frame normalized by the HMM models. After normalization, the exemplar is recognized by the neural net. The decision of the neural net is taken as the final result.

3.4. Modified Viterbi Algorithm

The straightforward application of Viterbi algorithm may not achieve full state segmentation in many situations. The Viterbi algorithm searches for the globally optimal state sequence. Often, the globally second best or the globally third best etc. sequences has the full state segmentation. The objective of the modified Viterbi algorithm is to find these sub-optimal sequences. There are two approaches -- parallel approach and serial approach. In the parallel approach, the trellis structure is extended to a third dimension. We call the third dimension the *dimension of choice*. Thus, all the nodes in the second choice plane represent the globally second best cost to reach that node and the transition from the previous layer. Similarly, all the nodes in the third choice plane represent the globally third best cost to reach that node and the transition from the previous layer; and so on. To track the globally optimal state sequence, all the terminal nodes in the first plane (choice 1) is used. The node that has the highest probability (or the lowest cost) is selected as the terminal state, and is used for retracing the state sequence. To track the globally second best state sequence, the terminal node used to find the globally optimal state sequence is replaced by its corresponding node in the second plane (choice 2). This node and all the terminal nodes in the first plane are used to track the globally second best state sequence. This idea is extended to track the globally third choice and so on. The compact mathematical representation of this algorithm is presented next.

Step 0: Storage:

t - time index

$\Psi_t(i,l)$, $1 \le t \le T, 1 \le i \le N, 1 \le l \le L$

- survivor terminating in state i at time t with choice l

$\delta_t(i,l)$, $1 \le t \le T, 1 \le i \le N, 1 \le l \le L$

- Max. prob. at state i at time t with choice l

Note that $\Psi_t(i,l)$ is a three-dimensional array, and each element of this array stores a two-dimensional data. $\delta_t(i,l)$ is a three-dimensional array, and each element of this array stores a one-dimensional data. For the Viterbi algorithm, we observe that both $\Psi_t(..)$ and $\delta_t(..)$ are 2-D arrays.

Step 1: Initialization:

$$\begin{aligned}
\delta_1(i,1) &= \pi_i b_i(o_1) && \text{for } 1 \le i \le N \\
\delta_1(i,l) &= 0 && \text{for } 1 \le i \le N, 2 \le l \le L \\
\Psi_1(i,l) &= (0,0) && \text{for } 1 \le i \le N, 1 \le l \le L
\end{aligned}$$

Step 2: Recursion: For $2 \le t \le T, 1 \le j \le N, 1 \le l \le L$,
compute $\delta_t(j,l)$ and $\Psi_t(j,l)$.

$$\delta_t(j,l) = (l-th) \max_{1 \le i \le N, 1 \le m \le l} [\delta_{i-1}(i,m)\, a_{ij}]\, b_j(o_t)$$

$$\Psi_t(j,l) = (i^*, m^*) = \arg(l-th) \max_{1 \le i \le N, 1 \le m \le l} [\delta_{t-1}(i,m)\, a_{ij}]$$

where $(c\text{-}th)$ max[.] denotes the c-th maximum.

Step 3: Termination: For $1 \le j \le N, 1 \le l \le L$

$$P^*(l) = (i-th) \max_{1 \le i \le N, 1 \le m \le l} [\delta_T(i,m)]$$

$$(i_T^*, l_T^*) = \arg(l-th) \max_{1 \le i \le N, 1 \le m \le l} [\delta_T(i,m)]$$

Step 4: Back-tracking: For $t = T-1, T-2, \ldots, 1$; and
$P^*(l), 1 \le l \le L;$

$$(i_t^*, l_t^*) = \Psi_{t+1}(i_{t+1}^*, l_{t+1}^*)$$

4. EXPERIMENTS

We have used the following data set for our experiments. We denote these signal classes as

Class A: Two broadband pulses of width 2.7ms and 3.7ms with variable spacing in between. Class B: Two broadband pulses of width 1.2ms and 1.5ms with variable spacing in between. Class C: Two broadband pulses of width 2ms and 3.1ms with variable spacing in between. Class D: Two broadband pulses of width 3.5ms and 2.7ms with variable spacing in between. Class E: 1.8-kHz tonal of variable duration, and 3.8 KHz tonal of variable duration with variable spacing in between the tonals. Class F: Three broadband pulses of width 3.1ms, 8ms and 6ms with variable spacing in between. Class F: Three broadband pulses of width 5ms, 7.1ms and 5.3ms with variable spacing in between. A typical example, one from each class, is shown in Fig. 2.

We have created 45 templates, i.e., exemplars, for each class, of which 23 are used as training templates and 22 as test templates. Each signal template contains 3072 data points. The sampling rate for the signal is 12,500 kHz. For this sampling rate, 3072 data points are enough to capture the essential characteristics of all the transient types. This 3072 point signal template is divided into 45 frames of 256 data points with an overlap of 192 points (75%) between two successive frames.

We have tried a different number of states for HMM, from N=2 to N=12, and a different number of nodes, from 10 to 30, in the hidden layer of the MLP-NN. Only the best results are reported in the paper and the accompanying table. Table 1 shows the number of errors in classifying the total 154 test exemplars using FFT power spectral features, QMF features and HMM, NN and integrated hybrid classifier. This dramatic improvement in classification accuracy (up to 8%) proves that the integrated hybrid classifier is indeed superior to both HMM and NN classifier considered individually.

5. ACKNOWLEDGMENT

The authors would like to thank ONR for supporting this research through the fellowship program administrated by ASEE and the independent research program administrated by Dr. Alan Gordon at NCCOSC, RDT&E Division.

REFERENCE

[1] J. Ghosh, L. M. Deuser, and S. D. Beck, ``A Neural Network based Hybrid System for Detection, Characterization and Classification of Short-Duration Oceanic Signals," IEEE Journal of Oceanic Engineering, Vol. 17, No. 4, pp. 351-363, 1992.

[2] A. Kundu, G. C. Chen, and C. E. Persons, "Transient Sonar Signal Classification Using HMM and Neural Net," IEEE Journal on Oceanic Engr., 19,1, pp. 87-99, Jan. 1994.

[3] P. Ramesh, S. Katagiri and C.-H. Lee, "A New Connected Word Recognition Algorithm Based on HMM/LVQ Segmentation and LVQ Classification", Proc. of ICASSP, vol. S, pp. 113-116, Toronto, Canada, 1991.

[4] R. P. Lippmann, "An Introduction to Computing With Neural Nets," ASSP Magazine, pp. 4-21, April, 1987.

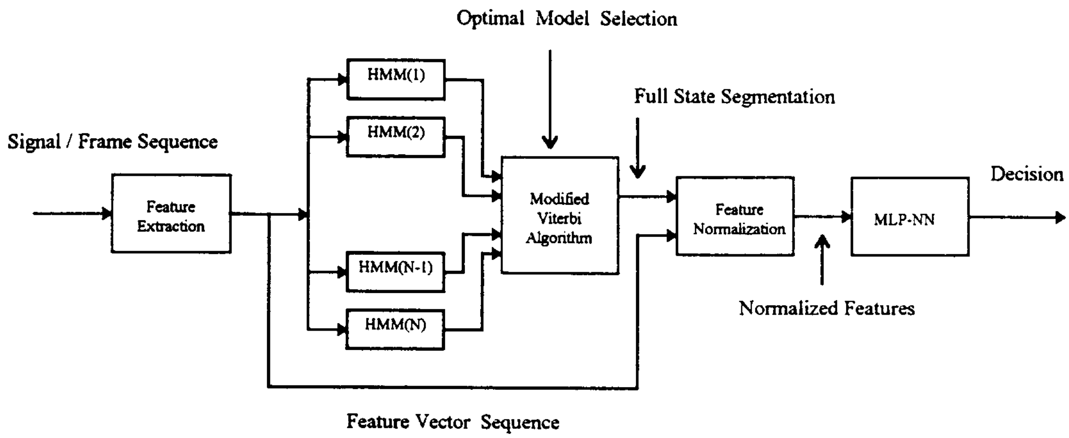

Figure 1. Integrated hybrid HMM/NN classification scheme.

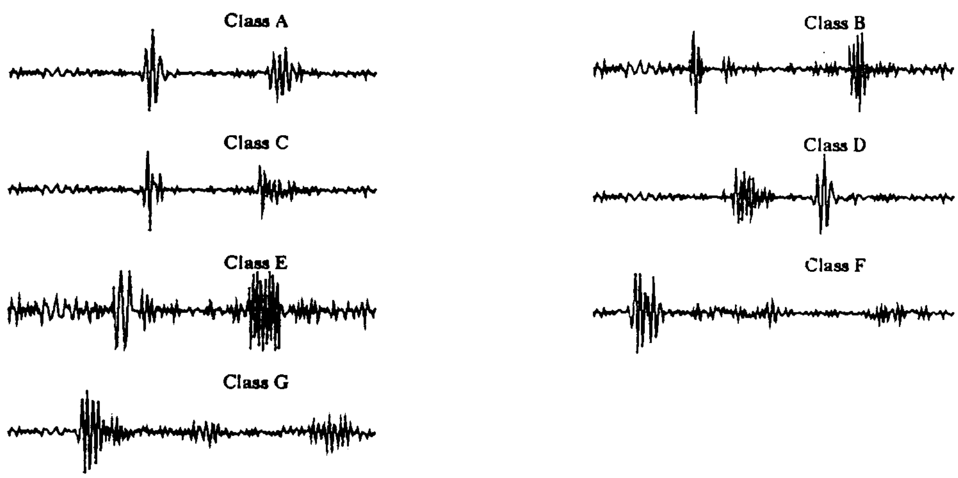

Figure 2. An example of the different classes of signals used in our experient.

Classifer/Feature	FFT Features	QMF Features
HMM	91 %	82 %
NN	88.3 %	72.7 %
HMM/NN	98.7 %	89.6 %

Table 1. Recognition performance of different classifiers.