Comparison the Performance of Hybrid HMM/MLP and RBF/LVQ ANN Models
Application for Speech and Medical Pattern Classification

Lilia Lazli[1], Abdennasser Chebira[2], Mounir Boukadoum[3] and Kurosh Madani[2]
[1]*Laboratory of Research in Computer Science (LRI/GRIA), Badji Mokhar University, B.P.12 Sidi Amar 23000 Annaba, Algeria*
[2] *Images, Signals and Intelligent Systems Laboratory (LISSI / EA 3956), PARIS XII University, Senart-Fontainebleau Institute of Technology, Bat. A, Av. Pierre Point, F-77127 Lieusaint, France*
[3]*Univesité du Québec A Montréal (UQAM), Montreal, Québec, Canada*

Keywords: Speech Recognition, Medical Diagnosis, Hybrid RBF/LVQ Model, Hybrid HMM/MLP Model.

Abstract: In the last several years, the hybrid models have become increasingly popular. We use involves multi-network RBF/LVQ structure and hybrid HMM/MLP model for speech recognition and medical diagnosis.

1 INTRODUCTION

The main difficulty in classification of speech and biomedical signals is related, on the one hand, to a large varity of such signals for a same diagnosis result by example the variation panel of corresponding medical signals could be very large, and on the other hand, to the close resemblance between such signals for two differents classification results. We propose two hybrid approaches for classification of electrical signals.

First, we have developed a serial multi-neural network approach that involves both Learning Vector Quantization (LVQ) and Radial Basis Function (RBF) Artificial Neural Networks (ANN). It is admitted that techniques based on single neural network show a number of attractive futurs to solves problems for which classical solutions have been limited, it is also admitted that a flat neural structure doesn`t represent the more appropriated way to approach "intelligent behavior".

We propose in second part a hybrid HMM/MLP model which makes it possible to join the discriminating capacities, resistance to the noise of MLP (Multi-Layer Perceptron) and the flexibilities of HMMs (Hidden Markov Model) in order to obtain better performances than traditional HMM.

2 DATABASES CONSTRUCTION

Three speech Data Bases (DB) and medical DB have been used in this work:

2.1 Speech Databases

1) The first one referred to as DB1, the isolated digits task has 13 words in the vocabulary: 1, 2, 3, 4, 5, 6, 7, 8, 9, zero, oh, yes and no, with a total of 3900 utterances.

2) The second DB2 contained about 50 speakers saying their last and first name, the city of birth and residence.

3) The DB3, contained the 13 control words (i.e. View/new, save/save as/save all), the used training set consists of 3900 sounds saying by 30 speakers.

2.2 Biomedical Database

The object is the classification of an electric signal coming from a medical test, the used signals are called Potentials Evoked Auditory (PEA) (Dujardin, 2006); (Lazli, 2007).

We choose 3 categories of patients (3 classes) according to the type of their trouble, we selected 213 signals, so that every signal contains 128 parameters. 92 belong to the Normal (N) class, 83, to the Endocochlear class (E) and 38 to the Retrocochlear class (R). The basis of training

Lazli L., Boukadoum M., Chebira A. and Madani K..
Comparison the Performance of Hybrid HMM/MLP and RBF/LVQ ANN Models - Application for Speech and Medical Pattern Classification.
DOI: 10.5220/0004133706590662
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ANNIIP-2012), pages 659-662
ISBN: 978-989-8565-21-1

contains 24 signals, of which 11 correspondent to the R class , 6 to the E class and 7 to the N class.

3 MULTI-NEURAL NETWORK BASED APPROACH

The approach we proposed to solve the problem is based on Multi-Neural Network (MNN) concept. A MNN could be seen as a neural structure including a set of similar neural networks (homogeneous MNN architecture) or a set of different neural nets (heterogeneous MNN architecture).

The serial homogeneous MNN is equivalent to a single neural network structure with a greater number of layers with different neuron activation functions. So the use of homogeneous MNN with a serial organization is here out of real interest. In the parallel homogeneous MNN configuration, each neural net operates as some "expert". So the interest of parallel homogeneous MNN appears when a decision stage, to process the results pointed out by the set of such "expert", is associated to such MNN structure becomes then a serial/parallel MNN, needing an optimization procedure to determine the number of neural nets to be used (Dujardin, 2006).

We propose an intermediary solution: a two stage serial heterogeneous MNN structure combining a RBF based classifier (operating as the first processing stage) with a LVQ based decision-classification stage see figure 1.

The RBF model we use is a weighted-RBF model but a standard one and so, it performs the feature space mapping associating a set of "categories" (in our case a category corresponds to a possible pathological class for example for medical DB) to a set of "areas" of the feature space. The LVQ neural model belongs to the the class of competitive neural network structure. It includes one hidden layer, called competitive layer. Even if the LVQ model has essentially been used for the classification tasks, the competitive nature of it`s learning strategy (based on winner takes all strategy), makes it usable as a decision-classification operator. On the other hand, the weighted nature of transfer functions between the input layer and the hidden one and between the hidden layer and the output one in this model allows non-linear approximation capability, making such neural net a function "approximation operator".

Taking into account the above analysis, the proposed serial MNN structure could be seen as a structure associating a neural decision operator to a neural classifier. Moreover, the proposed structure could also be seen as some global neural structure with two hidden layers. So, the association of two neural models improves the global order of the non linear approximation capability of such global neural operator, comparing to each single neural structure constituting the MNN system. This technique allows to fill in the gap induced by the RBF ANN, and thus, to refine the classification.

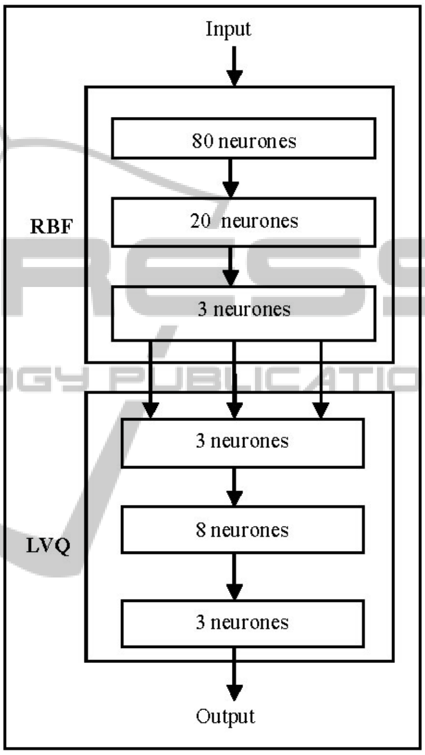

Figure 1: Serial Multi-Neural Network based structure.

4 HYBRID HMM-ANN MODELS

We describe the theoretical formulation of our hybrid HMM/ANN model, an approach for the training and estimation of posterior probabilities using a recursive algorithm that is reminiscent of the EM (Expectation Maximization) algorithm for the estimation of data likelihoods. The method is developped in the context of a statistical model for transition-based electrical signal recognition using ANN to generate probabilities for HMM. In the new approach, we use local conditional posterior probabilities of transitions to estimate global posterior probabilities of instance sequences given acoustic data.

4.1 Estimating HMM Likelihoods with ANN

ANN can be used to classify pattern classes (Lazli and Sellami, 2003). For statistical recognition systems, the role of the local estimator is to approximate probabilities or Probability Density Functions (PDF). Practically, given the basic HMM equations, we would like to estimate something like $p(x_n \backslash q_k)$, is the value of the PDF of the observed data vector given the hypothesized HMM state. The ANN can be trained to produce the posterior probability $p(q_k \backslash x_n)$ of the HMM state give the acoustic data (figure 2). This can be converted to emission PDF values using Bayes'rule. Since the network outputs approximate Bayesian probabilities, $g_k(x_n,\Theta)$ is an estimate of:

$$p(q_k \backslash x_n) = \frac{p(x_n \backslash q_k)\, p(q_k)}{p(x_n)} \qquad (1)$$

Which implicitly contains a priori class probability $p(q_k)$. It is thus possible to vary the class priors during classification without retraining, since these probabilities occur only as multiplicative terms in producing the network outputs. As a result, class probabilities can be adjusted during use of a classifier to compensate for training data with class probabilities that are not representative of actual use or test conditions.

Thus, scaled likelihoods $p(x_n \backslash q_k)$ for use as emission probabilities in standard HMM can be obtained by dividing the network outputs $g_k(x_n)$ by the training set, which gives us an estimate of:

$$\frac{p(x_n \backslash q_k)}{p(x_n)} \qquad (2)$$

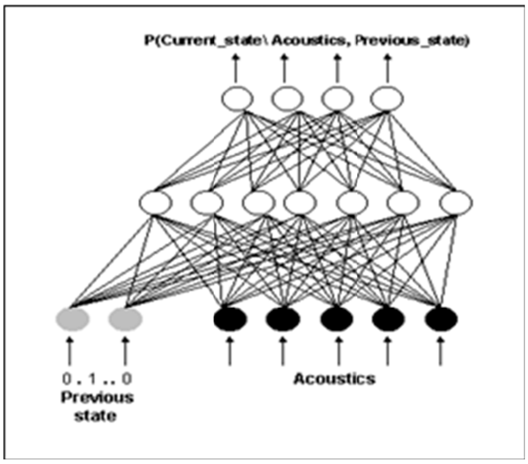

Figure 2: An MLP that estimates local conditional transition probabilities.

During recognition, the seating factor $p(x_n)$ is a constant for all classes and will not change the classification. It could be argued that, when dividing by the priors, were using a scaled likelihood, which is no longer a discriminant criterion. However, since the discriminant training has affected the parametric optimization for the system that is used during recognition. Thus, this permit uses of the standard HMM formalism, while taking advantage of ANN characteristics.

5 CASE STUDY AND EXPERIMNTAL RESULTS

5.1 Comparison Study with Single RBF and LVQ Approaches

The learning DB has successfully been learnt. The **RBF** network well classifies **62,3**% of the full DB, with a rate of correct classification of: 61% for the R class, 58% for the E class, 68% for the N class.

For the tree speech DBs, a rate of correct classification as follows: **51**% for the DB1, **52**% for the DB2, **63**% for the DB3.

The **LVQ** network well classifies **62**% of the full medical DB with a rate of correct classification of: 72% for the R class, 57% for the E class, 57% for the N class. For the tree speech DBs, a rate of correct classification as follows: **65**% for the DB1, **61**% for the DB2, **68**% for the DB3.

5.2 Results Relative to RBF/LVQ based Multi-neural Network Approach

Concerning the **RBF** ANN, by example for the biomedical DB, the number of input neurons (88) corresponding to the number of components of the input vectors, the output layer contain 3 neurons. The number of neurons of the hidden layer is 20 neurons.

For the **LVQ** ANN, the number of input cells is equal to the number of output cells of the RBF ANN. The output layer of the LVQ ANN contains as many neurons as classes (3). The number of neurons in hidden layer is 8 neurons has been determined by considering the number of subclasses we can count into the 3 classes.

The learning DB has successfully been learnt. All of the learnt vectors are well classified in the generalization phase. We can see that this network well classifies 65% of the full DB, with a rate of correct classification of: 71% for the R class, 55%

for the E class, 69% for the N class.

For the speech DBs, a rate of correct classification as follows: 79% for the DB1, 86% for the DB2, 72% for the DB3.

Comparing to the two single approaches, with our proposed MNN technique, we obtain globally better results than the single RBF or LVQ ANN approach.

5.3 Results Relative to Discrete HMM and Hybrid HMM/MLP Approach

Further assume that for each class in the vocabulary we have a training set of k occurrences (instances) of each class where each instance of the categories constitutes an observation sequence.

a. Discrete HMM

For speech DBs, 10-state, strictly left-to-right, discrete HMM were used to model each basic unit (words). In this case, the acoustic feature were quantizied into 4 independent codebooks according to the KM algorithm: 128 clusters for the J RASTA-PLP coefficients, 128 for the Δ J RASTA-PLP coefficients, 32 clusters for Δ energy, 32 clusters for $\Delta\Delta$ energy.

For the PEA signals, 5-state, strictly left-to-right, discrete HMM were used. Table 1 gives the results of this experiment.

Table 1: Discrete HMM results.

	BDB	SDB1	SDB2	SDB3
Rate%	84	87	90	76

b. Discrete MLP with entries provided by the FCM Algorithm

We use an hybrid HMM/MLP model using in entry of the ANN an acoustic vector composed of real values which were obtained by applying the FCM algorithm (Lazli and Sellami, 2003) with 2880 real components corresponding to the various membership degrees of the acoustic vectors to the classes of the "code-book". We reported the values used for SDB2.

Table 2: Hybrid HMM/MLP results.

	BDB	SDB1	SDB2	SDB3
Rates %	94	97	97	83

10-state, strictly left-to-right, word HMM with emission probabilities computed from an MLP with 9 frames of quantizied acoustic vectors at the input. Thus a MLP with only one hidden layer including 2880 neurons at the entry, 30 neurons for the hidden layer and 10 output neurons was trained.

For the PEA signals, a MLP with 64 neurons at the entry, 18 neurons for the hidden layer and 5 output neurons was trained. Table 2 gives the results.

6 CONCLUSIONS

In this paper, the association of RBF and LVQ neural models improves the global order of the non linear approximation capability of such global neural operator, comparing to each single neural structure constituting the MNN system.

For the second hybrid HMM/MLP model, a recognition tasks show an increase in the estimates of the posterior probabilities of the correct class after training.

From the effectiveness, it seems that the hybrid HMM/MLP model are more powerful than multi-network RBF/LVQ structure.

REFERENCES

L. Lazli, M. Sellami. "Connectionist Probability Estimators in HMM Speech Recognition using Fuzzy Logic". *MLDM 2003: the 3rd international conference on Machine Learning & Data Mining in pattern recognition*, LNAI 2734, Springer-verlag, pp.379-388, July 5- 7, Leipzig, Germany, 2003.

A-S. Dujardin. "Pertinence d'une approche hybride multi-neuronale dans la résolution de problèmes liés au diagnostic industrièle ou médical". *Internal report,* I2S laboratory, IUT of "Sénart Fontainebleau", University of Paris XII, Avenue Pierre Point, 77127 Lieusant, France, 2006.

L. Lazli, A-N. Chebira, K. Madani. "Hidden Markov Models for Complex Pattern Classification". *Ninth International Conference on Pattern Recognition and Information Processing, PRIP'07* may 22-24, Minsk, Belarus, 2007. http://uiip.basnet. by/conf/prip2007/ prip2007.php-id=200.htm