ISSN:0975-9646
Geetika Munjal / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (4) , 2011, 1555-1558

ANN Paradigms for Audio Pattern Recognition

Geetika Munjal
ITM University,Sector 23A Gurgaon, India

***Abstract*- Pattern Recognition is the process to classify data or patterns based on either a priori knowledge or on statistical information extracted from the patterns. An audio pattern recognition problem is based on speech patterns spoken, which can be interpreted as speaker dependent or speaker independent. Artificial Neural Network (ANN) is information processing machine learning model, inspired by biological neural systems. ANN has a potential of massive computing, online adaptation and learning abilities. Neural network consists of many simple processing elements joined by weighted connection paths. A neural net produces an output signal in response to an input pattern; the output is determined by value of weights. This paper discusses various neural network paradigms for audio pattern recognition and based on the study a new paradigm for audio pattern recognition is suggested.**

Keywords: Artificial Neural Network, Pattern Recognition Self Organizing Maps, Learning Vector Quantization, Multilayer Perceptron, Learning, Divide and Conquer, Mel Frequency Cepstrum Coefficient, Linear predictive coding

I. INTRODUCTION

Pattern recognition is a process to determine whether an input pattern is a member of a particular class or not. It is problem in which similar patterns are grouped together the grouping are then defined as classes. Speech Pattern recognition can be categorized into speaker dependent (who is speaking) and speaker independent (what is being spoken?)[1]. In both ways speech is the major component. Speaker dependent system is for speaker recognition system which can be further divided into speaker identification (Who is speaking?) and speaker verification (Is the speaker in the same authentic person?). The speaker recognition process can be divided mainly in two modules i.e. feature extraction and feature mapping. Before feature extraction some preprocessing like framing and windowing is done[6]. In feature extraction some useful data or training data can be obtained using LPC or Cepstrum analysis other methods for feature extraction is MFCC. Then in matching phase the testing data is compared with the training data to obtain the result. Various methods of speaker recognition exists like distortion based, HMM based, ANN [2], DTW. Artificial neural network is a connectionist model, which has a vast application in pattern recognition. Neural network have a potential of massive computing and learning abilities, it provides a new approach to speaker recognition. The factors that influence neural network for speaker recognition are the input pattern, the scale of neural network, type of speaker recognition i.e. text dependent or text independent. Neural networks lack centralized control, in the classical sense, since all the interconnected processing elements change or "adapt" simultaneously with the flow of information and adaptive rules. One of the original aims of artificial neural networks (ANN) was to understand and shape the functional characteristics and computational properties of the brain. Types of learning in ANN are supervised and unsupervised. In supervised the network is trained using a set of input-output pairs. The goal is to 'teach' the network to identify the given input with the desired output. For each example in the training set, the network receives an input and produces an actual output. After each trial, the network compares the actual with the desired output and corrects any difference by slightly adjusting all the weights in the network until the output produced is similar enough to the desired output, or the network cannot improve its performance any further. In unsupervised learning the network is trained using input signals only. In response, the network organizes internally to produce outputs that are consistent with a particular stimulus or group of similar stimuli. Inputs form clusters in the input space, where each cluster represents a set of elements of the real world with some common features. Various Neural Networks like perceptron based, Competitive Learning based, Radial Basis Function based support audio pattern recognition problems. All the ANN are discussed in the next sections.

II. PERCEPTRON BASED AUDIO PATTERN RECOGNITION

The Perceptron is the basic processing element and simplest learning machine that is based on supervisory training[4]. Perceptron algorithm is an implementation of gradient decent method, according to which the mean squared error E (w) has associated with it a gradient E. The vector E point in the direction in which E (w) will decrease at the fastest possible rate and weights are updated with equation

$$w(k+1) = w(k) - c(E) \quad \text{……..(1)}$$

where c is suitable constant, and the activation

$$y = f(\sum wj * xj + w\,o) \quad \text{.. (2)}$$

In a recognition with K classes there are K perceptions where $y^t_{i=1}$ if $x \wedge t \in Ci$ if and $y^t_i = 0$ otherwise. Perceptron initially seemed promising, but later on it was realized that single layer perceptron are capable of learning only linearly separable patterns this lead to research of MLP that have a grater processing power than perceptron with one layer In MLP there are one or more hidden layers depending on arbitrary pattern, binary pattern or non-binary pattern. From

network complexity performance and implementation consideration a larger no. of hidden layer with corresponding increase in the number of hidden units and connections may be required In MLP sigmoid function is used for activation represented by

$$h_k = sigmoid\left(w_k^T x + w_{ko} = \frac{1}{1+\exp(-\sum_{j=1}^{d} w_{kj}x_j + w_{ko}} \right) \quad \ldots\ldots(3)$$

The algorithm starts with initial weights which are randomly assigned and updated based on derivative of errors

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial_y} \frac{\partial_y}{\partial h_k} \frac{\partial h_k}{\partial w_{kj}} \quad \ldots\ldots..(4)$$

function here back propagation algorithm provides a way to calculate the gradient of error efficiently. The error of the initial computation is forward pass is propagated backward from the output units, layer by layer justifying the name back propagation.

The BPA is simplest, general and widely used for training the Multilayer feedforward network. However there is no guarantee of convergence to the right solution thus lack of convergence is severe drawback of back propagation especially when different classes of pattern are close to each other in multidimensional feature space.

In the recognition experiment for music, the accuracy achieved for various types of data set in which the success rate achieved is 91% to 96.7% by MLP network[6]. Features given to the neural network here plays a major role in recognition accuracy.

III. COMPETITIVE LEARNING BASED AUDIO PATTERN RECOGNITION

In a pattern recognition problem the network gives response in two or more classes, i.e. several neurons are responding to the input vector[4]. In such situation where we know that only one of the several neurons should respond we can include additional structure in the network so that the net is forced to make a decision as to which unit will respond .The mechanism by which this is achieved is called competition. The most extreme form of competition among group of neurons is called winner takes all**.** As the name suggest, only one neuron in the competing group will have a nonzero output. Output signal will be produced when the competition is completed. In computer simulations of these nets, full neural implementation of the algorithms is not primary importance, it is easy to replace the iterative competition phase of the process with simple search for neuron with the largest input chosen as the winner. The neural nets based on this type of learning are self organizing maps (SOM) developed by kohonen and learning vector quantization (LVQ)[4]. SOM group the input data in to clusters, and it is a common use for unsupervised learning. LVQ classifies an input vector by assigning it same class as the output unit that has its weight vector closest to the input vector. In original LVQ [4] only the reference vector updated that is closest to the input vectors updated. The direction it is moved depends on whether the winning neuron reference vector belongs to same class as the input vector. This algorithm can be improved if two vectors i.e. winner and runner up learn if (i) they both belong to two different classes ,(ii) the input vector belongs to the same class as the runner up.

The LVQ performs very well if suitable initialization of weights is done. Training an LVQ is accomplished by presenting input vectors and adjusting the location of hidden units based on their proximity to the input vector. The nearest hidden units based is moved a distance proportional to the learning rate. The hidden layer weights are trained in this manner for an arbitrary number of iterations, usually with learning rate decreasing as the training progresses. The objective is to place the hidden units so as to cover the decision regions of the training set. LVQ have been found to perform well in pattern recognition but processing required for input classification may be larger since more hidden units are often required. The LVQ is based on competitive learning so the stability of the clusters is not guaranteed, it can be achieved by gradually reducing the learning rate to zero but learning rate should be increased to learn new patterns. Self organizing map is a method of machine learning where a model is fit to observations. It is distinguished from supervised learning by the fact that there is no priori output. In unsupervised learning, a data set of input objects is gathered. Unsupervised learning then typically treats input objects as a set of random variables. LVQ for pattern recognition it has scored varying accuracy from 66.7% to 98% based on the input pattern quality[1], and depending upon the number of hidden units (in this case it is 2) it gives an accuracy of 88.8% and 89.2%.if a multilayer version of LVQ is used then the recognition rate achieved is 65.5 with 100 hidden neurons, here the computational complexity is very high than classical LVQ[1].

The self organizing map is a subtype of artificial neural networks. It is trained using unsupervised learning to produce low dimensional representation of the training samples while preserving the topological properties of the input space. This makes SOM especially good for visualizing high-dimensional data. The training utilizes competitive learning. When a training sample is given to the network, its Euclidean distance to all weight vectors is computed. The neuron with weight vector most similar to the input is called the Best Matching Unit (BMU). The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector. The magnitude of the change decreases with time and

is smaller for neurons physically far away from the BMU. SOM [5] operates in two modes training process in which the map organizes itself and mapping process in which a new input vector may quickly be given a location on the map.
Competitive learning algorithm is used in training of SOM classifier. Because of unsupervised learning schema, SOM classifier is organized them without any external impact. Therefore SOM classifiers converge to a global solution more quickly than supervised learning classifiers SOM network have only one layer and all neuron are fully connected to the inputs [3]. The outputs of SOM classifier show the indices of activated neuron for a particular input pattern [5]. When a SOM classifier is tested, we can only observe which output neuron is activated for the input pattern. In SOM the weights are true representative of probability p(x) of the inputs used for learning i.e. the weights are distributed evenly for uniform input density distribution.
Multilayer SOM is a complex network with many different functional units the signal flows through two paths forward path and backward path this type of SOM can be helpful in recognition of high dimensional data like images or visual recognition systems.
SOM [5] perform much better than MLP in case of speech pattern recognition if it is two dimensional with 5000 epochs with a size of 1000*1000 the accuracy achieved is 97.4% .

IV. PROPOSED PARADIGM

In this paper a two tier architecture of Kohnen's SOM and LVQ is proposed for feature mapping/recognition. Kohnen's self organizing map is fed with the input vector or extracted MFCC as SOM classifier with MFCC gives sufficient results, SOM uses unsupervised learning to modify the internal state of the network and to model the features found in the training dataset. The map is automatically organized by a cyclic process of comparing the input patterns to the vectors at each node. The node vector with which the input matches is selectively optimized set of nodes; this method represents the prototype of the patterns. The output obtained is further fine-tuned using LVQ technique. The model is named as EX-LVQ

EX-LVQ algorithm

Step 0 Initialize weights w_{ij} either by setting topological parameters or setting learning parameters
Step1 While stopping condition is false do step 2-8
Step2 for each input vector exclude steps 3-5
Step3 for each j, compute:

$$D(j) = \sum (w_{ij} - x_i)\text{^}2$$

Step4 Find index *j* such that *D(j)* is minimum
Step5 For all units *j* within a specified neighborhood of *j* and all for all *i find* w_{ij} :

$$w_{ij}(new) = w_{ij}(old) + a[x_i - w_{ij}(old)]$$

Step 6 update learning rate.
Step 7 Reduce radius topological neighborhoods at specified times.
Step 8 Test stopping condition for SOM algorithm
Step9 $w_{ij}(new1) = w_{ij}(new) + a[x_i - w_{ij}(new)]$
Step 10 For each training input vector x, do steps 11,12
Step 11 find j so that $|x - w|$ is minimum
Step 12 again update weight as follows
If $T = Cj$, then

$$w_{ij}(new1) = w_{ij}(new) + a[x_i - w_{ij}(new)]$$

If $T \neq C_j$, then

$$w_{ij}(new1) = w_{ij}(new) - a[x_i - w_{ij}(new)]$$

Step 13 Reduce learning rate
Step 14 Test stopping condition

The stopping condition may specify a fixed number of iterations or the learning rate reaching a sufficiently small value.
The nomenclature used in the algorithm is as follows
x training vector
T correct category or class of training vector
w_{ij} weight vector for jth output neuron unit$(w_{1j}....w_{ij}.....w_{nj})$
Cj class represented by j_{th} output unit.
|x-w| Euclidean distance between input vector and j_{th} output unit.

The w_{ij}(new) is the initialized weight vector for kohnen's SOM and the w_{ij}(new1) is the updated weights obtained and fed into LVQ. The SOM has classified the input patterns and updated the initial weights. The input vector fed in LVQ is the classified patterns obtained from SOM. The initial weight vectors are randomly selected from the input vector, The Euclidean distance D is to find the winner unit of all SOM units which is based on Best Matching Unit (BMU) where the distance between the input vector and weight vector should be minimum. The SOM preserves the local topology of the input space in the low dimensional display as faithfully as possible which means the inputs are mapped to nearby SOM units. In SOM the first training epochs with large neighborhoods draw automatically most units near the important areas and enable their later use. The SOM is not intended for optimal recognition but classification accuracy of SOM codebook can be improved by LVQ methods [3]. The LVQ aims at defining the decision surfaces between the competing classes thus leading to minimization of recognition error, the lowest error rate is obtained by

concentrating on the actual discrimination between the classes. The method based on neural networks may outperform other methods in tough problems, where the prior knowledge cannot help much in the recognition and the system characteristics must be learned automatically from the data[3]. The input vectors fed in the SOM are extracted features obtained with the help of MFCC, w_{ij}(new) is chosen randomly from the input patterns, later on w_{ij}(new) is updated to w_{ij}(new1) and fed to LVQ.

V. RESULTS AND CONCLUSION

The performance of ANN for any audio pattern recognition will depend on many factors including the quality of input pattern fed in the neural network, the quantity of input pattern, the scale of neural network including the number of hidden layers and number of hidden units in each layer. MLP can give good results in pattern recognition depending of the size of the network and the quality of input patterns, but it is based on back propagation network so have another drawback of lack of convergence. Classical LVQ gives less accurate results than MLP but can perform better if appropriate number of hidden neurons is chosen and input patterns incase of LVQ2 and LVQ3. SOM performs well in audio pattern recognition. Their recognition accuracy can be multiplied if nodes are further fine tuned using supervised learning; its performance also depends on the dimension and epochs.

The recognition accuracy and factors effecting recognition accuracy of three ANN is shown in table

Artificial Neural Network	Recognition Accuracy (Percentage)	Factors effecting recognition accuracy
LVQ	66.7-98	1. Pattern quality 2. Number of hidden units
SOM	97.4	1. Number of layers 2. Good for high dimensional data
Multilayer Perceptron	91-96.7	1. Input feature or pattern 2. Network size

The pattern recognition quality can be evaluated depending on various metrics like pattern similarity measure, which can be calculated by dot products of two patterns, divided by the length of two patterns. In the proposed architecture, Advantages of SOM and LVQ are merged in a single paradigm thus combining the benefits both supervised and unsupervised learning.

VI. REFERENCES

1. Brian J. Love, Jennifer Vining, Xuening Sun *"Automatic Speaker Recognition Using Neural Networks"* , EE371D spring 2004.
1. G. Rigoll *"Speech recognition experiments with a new multilayer LVQ network*" EUROSPEECH-95 Spain, pp 2167-2170, 1995
2. H. Hattori *"Text independent speaker recognition using Neural Network"* IEEE Sanfrancisco, USA, vol 2, pp 153-156, 1992.
3. K.V Prema and S. Reddy. *"Two-tier architecture for unconstrained handwritten Character recognition*" IAS (Indian Academy of Science), Bangalore vol 27, pp 585-594, Oct 2002.
4. L. Fausset *"Fundamentals of Neural Network Architectures, Algorithms and applications"* 3rd Edition, Pearson education publishers, pp 174-206.
5. M. Inal and S. Yuvaz *"Self organizing map and associative memory model Hybrid classifier for speaker recognition*" IEEE, pp 8129-8132, 2002.
6. P. Scott *"Music Classification using Neural Networks***"** EE73B project report, pp 1-5, 2001.
7. R. Wouhaybi and Adnan.M *"Comparison of Neural Networks for speaker Recognition"* IEEE, vol 1, pp. 125-129, 1999.