COMPARATIVE STUDY OF ANN FOR PATTERN CLASSIFICATION

GEETIKA MUNJAL and SUNINT KAUR
Dept. of Computer Science and Engg.
Guru Nanak Dev Engg College, Ludhiana 141006

Abstract
Pattern classification is a process to determine whether an input pattern is or is not a member of a particular class. It is problem in which similar patterns are grouped together the grouping are then defined as classes. Pattern classification is one type of pattern recognition which has a lot of applications including Finger print classification, handwritten character recognition, speaker recognition. Artificial Neural network (ANN) is machine learning model which are information processing systems, inspired by biological neural systems. ANN have a potential of massive computation, 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 makes a comparative study of various neural networks for pattern classification. The neural networks discussed in this paper are Perceptron based, VoD based SOM and RBF network

Keywords: *Artificial Neural Network, Pattern Classification Self Organizing Maps, Learning vector Quantization, Multilayer perceptron, Voronoi Vector, Learning, Divide and conquer*

1. Introduction

Pattern classification algorithms receive patterns as input and try to understand if these patterns own some specific property or class. The pattern is generally described as a feature vector in which each component is a vector which is obtained by some techniques that gives some measurable values to every pattern. The pattern classification[5] has a lot of application like Fingerprint Classification Speaker Identification Medical Applications The[7] classifier operate steps in two **training phase** during which it is provided with specific knowledge on the considered application domain, using information about a representative set of samples (training set) described according to the considered description scheme and **operative phase** in which the classifier is first given the description of a sample to be recognized and then assigns it to a class on the basis of the experience acquired in the training phase. Ann provides a great flexibility in learning different real life problems and pattern classification is one of them. The classification or description scheme is usually based on the availability of a set of patterns that have already been classified or described. This set of patterns is termed the training set and the resulting learning strategy is characterized as supervised learning. Learning can also be unsupervised, in the sense that the system is not given an a priori labeling of patterns, instead it establishes the classes itself based on the statistical regularities of the patterns.

The classification or description scheme usually uses one of the following approaches: statistical (or decision theoretic), syntactic (or structural). Statistical pattern recognition is based on statistical characterizations of patterns, assuming that the patterns are generated by a probabilistic system. Structural pattern recognition is based on the structural interrelationships of features. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. The real life application of ANN includes Function approximation, or regression analysis, including time series prediction and modeling, Classification, including pattern and sequence recognition, sequential decision making and Data processing.

2. Perceptron based classification

The Perceptron [8, 4] is the basic processing element and simplest learning machine that is based on supervisory training .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 eq. $w(k+1) = w(k) - c(E)$ -------------------(1)

where c is suitable constant. And the activation $y = f(\sum wj * xj + wo)$. (2)

In a classification 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 which have a grater processing power than perceptron with one layer

2.2. Multilayer Perceptron (MLP)

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

$(w^t_k x + w_{k0}) = 1/(1+\exp(-\Sigma^d_{j=1} w_{kj} x_{j+} w_{ko})))$ --(3)

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

$\partial E/\partial w_{kj} = (\partial E/\partial y)(\partial y/\partial h_k)(\partial h_k/\partial w_{kj})$------(4)

here. Back propagation algorithm provide 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.

3. Classification based on VoD

VoD[5] based classification has excellent discrimination capabilities in situation where patterns are close to each other in multidimensional space and are also able to perform better in noisy environment even if the best pattern deviate from the exampler .A VoD of a set of specified points sometimes referred to as sites space that assign a surrounding region or voronoi cell (representable of the intersection of afinite number of closed half space) of nearby points to each of the pattern sites. The voronoi cell around a chosen site describes a region that contains interior points that are nearer to the site than any other site.

The VoD based classification if analyzed on the basis of size and complexity then they are less complex and more robust. LVQ i.e. Learning Vector quantization is based on VoD in which each output neuron represents a particular class or category the weight vector for an output is often referred to as voronoi or feature space for that class that the unit represents .During training the output neuron are positioned by adjusting the weight through supervised training .It is assumed that a set of training patterns with known classifications is provided, along with an initial distribution of reference vectors each of which represents a known class. 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 classification but processing required for input classification may be larger since more hidden units are often required.

In VoD based learning if some additional learning growth algorithm (divide and conquer, The Upstart algorithm, Tilling algorithm) are included it will lead to adaptive network it will lead to adaptive network [7]. These algorithms are based on decomposition and subgoaling, based on this principle a problem is decomposed into subgoals so that each subgoal can be learned quickly using these subnetworks, and achieves the global goal by putting learnable by subnetworks. The basic strategy applied to all growth algorithm is divide and conquer i.e. divide the original training set into smaller subset and test to see whether each subset may be learned separately .If a subset still cannot be learned within a specified time, divide it up again, Eventually in the worst case a subset contains only two exemplars, each belonging to a different class is obtained. These growth algorithms may lead to a large network leading to poor generalization. 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. For that adaptive network were introduced which includes ART1 (adaptive resonance theory).

4. SOM based Pattern classification Unsupervised learning [7]

It 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. The **self-organizing map (SOM)** 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 operates in two modes Training process in which the map organizes itself, using a competitive process 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. The outputs of SOM classifier show the indices of activated neuron for a particular input pattern. 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 classification of high dimensional data like images or visual recognition systems.

5. Radial Basis Function (RBF) based Classification

RBF is a two layer network in which each hidden unit implements a radial activated function, RBF implements Bayesian rule (which is based on probabilistic Bayes decision theory) and model any continuous input and output mapping. The output unit implements a radial a weighted sum of hidden nodes; if the hidden nodes have fewer degrees of freedom per node then this network can perform better. RBF networks are generally used in supervised applications. In the training algorithm of RBF each sample is assigned some basis function but it makes it more memory consuming so hidden unit weights are chosen randomly and output weights are calculated .Training is accomplished first at hidden layer by unsupervised learning and then at the output layer by supervised learning. At the hidden layer the centroids are calculated by k-means clustering or as said earlier no. of exemplars equal to no. of hidden units. The hidden units are trimmed as desired adjusting the number of hidden unit as desired. Learning in the output layer follows the determination of parameters in the activation function either by Least Mean Square algorithm (learning rule based on minimization of squared error for each training pattern also called Widrow-Hoff or delta rule). After this initial training of the network, the network parameters in the hidden layer may fine tune by applying supervised learning to both the hidden and the output layers simultaneously. RBF network can provide arbitrarily good approximations depending on the size of hidden layer .RBF have fast convergence rate (The configuration of net stops changing if the weights reach equilibrium) net and have simple network structure with easier control over network performance.

In pattern classification the input represents the feature entries, the output correspond to a class and the hidden neuron correspond to subclasses. In RBF Gaussian function is preferred as activation function denoted by Φ .The output layer implements a weighted sum of hidden units

$$\psi(X) = \sum_{j=1}^{L} \lambda jk \phi j(X) \qquad (5)$$

where λ_{jk} are the output weights corresponding connection, between a hidden unit and output unit. The weights represent the contribution of the output unit. In pattern classification problem the out of RBF lies between [0, 1].The input vector is expanded into hidden units space providing linear seperability.

6. Discussion

In the classification 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 .Features given to the neural network here plays a major role in classification accuracy. Talking about LVQ for pattern recognition it has scored varying accuracy from 66.7% to 98% based on the input pattern quality, 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. SOM [6] perform much better than MLP in case of speech pattern classification if it is two dimensional with 5000 epochs with a size of 1000*1000 the accuracy achieved is 97.4% .The performance of RBF network can be compared to that of MLP and LVQ ,RBF partitions the feature space better than MLP. Performance of RBF network increases with the number of RBF function but at some values the effect is negligible.

7. Deduced Results

The performance of ANN for any application 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 no. of hidden units in each layer. MLP can give good results in pattern classification 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 classification. Their classification accuracy can be multiplied if nodes are further fine tuned using supervised learning; its performance also depends on the dimension and epochs. RBF networks uses much more general and versatile than competitive learning. The performance of RBF depends on the hidden units which generally represent the class of pattern. The pattern classification 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.

8. Conclusion

In this paper a comparative study of neural network for pattern classification is done, various neural network for pattern classification are discussed including perceptron based learning, competitive based learning , unsupervised learning all these neural network can perform pattern classification very well , but the performance of all these networks depends on various factors and the performance can further be improved by taking care of certain things in every neural network as discussed in deduced results.

References

1. *Adrain G.bors "Introduction to Radial Basis Function Network" university of York ,UK.*
2. *Brad A. Hawickhorst and Stephen A. Zahorian "A Comparison Of three Neural Network Architecture for Automatic Speech Recognition ".Department of Electrical and Computer Engineering "Old Dominion University, Norfolk, VA .*
3. *Fan Sun ,Bibo Li Huisheng Chi,"Some key factors in Speaker Recognition using Neural Network" National Labortary Machine Perception ,Peking University*
4. *G.Rigoli "Speech Recognition Experiments with a New Multilayer LVQ Network" Gehard-Mercator –University Duisburg, Germany*
5. *K V Prema and N V Subba Reddy "Two-tier architecture for unconstrained handwritten character recognition" Deprtment of CSE and Engg ,Manipal Institute of technology pp 585-594 oct-2002*
6. *Laurene Fausett'Fundamentals of Neural Netwoks Architecture,Algorithms and applications*
7. *Melih INAL ,Yavuz selim FATIHOGLU "Self organizing man and associative Memory Model" Kocaeli University ,technical Education Faculty,Electronins and computer education Izmit/Kocaeli/TURKEY Sept 2002*
8. *N.K Bose and P.Liang "Neural Network Fundamentals with graph,Algorithms and applications" Tata McGraw Hill*
9. *P. Geittner "Radial Basis Network for speech Classification"Philips Research laboratory Fedral Repblic of Germany 1988 , 437-439*
10. *Paul Scott"Music Classification using Neural Network" Spring 2001*
11. *Rita H Wauhaybi and Mohamad Adnan Ai-Alaoui "Comparison of Neural Networks For Speaker Recognition "*
12. *S.Quamour-Sayoud M.Boudraa "Application of MLVQ1 in Speaker identification" Institute of Electronique*
13. *SE Fredrickson and L.Tarassenko "Text independent speaker recognition using neural network techniques" Qxford University, UK.*