International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 05, May 2016
www.ijirse.com
ISSN: 2454-9665

NEURAL NETWORK AND STATISTICAL PATTERN RECOGNITION TECHNIQUES: REVIEW PAPER

Ritika Dewan[1] , Lakshit[2]

[1]*ECE, BMIET/DCRUST , (India)*

[2]*ECE, SGI/KUK , (India)*

ABSTRACT

Pattern Recognition is an attractive field of machine learning from past few years. Primary goal of handwritten numeral statistical pattern recognition techniques is to efficiently differentiate between statistical and neural pattern recognition technique and tried to realize revealing new techniques for far better results than statistical techniques. Among various frameworks in which pattern recognition has been traditionally formulated, statistical techniques intensively used in practice. In comparison , Discriminant Analysis (DA) and Principal Component Analysis (PCA) we are using them for pattern analysis or recognition, which are a statistical technique. Discriminant analysis engrosses problem regarding dimensional data and smaller size of samples. To abolish these problems, pattern recognition task is implemented using Generalized Regression Neural Network (GRNN) and Back propagation neural network (BPNN) techniques. Pattern Recognition task is conceded on data base of face images of 400 people. Neural network proved results far better than statistical methods.

Keywords: Pattern Recognition, Discriminant Analysis, Principal Component Analysis, Generalized Regression Neural Network, Back-Propagation Neural Network.

I INTRODUCTION

Most people and children recognize digits and letters whether they are small characters, large characters, handwritten, machine printable, rotated, straight and so on [1]. These patterns are recognized by the young very easily. Pattern recognition techniques are used to process data and also for decision making. As best pattern recognizers are humans. We want this ability in machines to do the same work as that of humans by using artificial intelligence. Pattern Recognition is actually a kind of study of machines that how a machine teach and learn to distinguish pattern of interest from their background, and make sound and reasonable decision about categories of patterns [4] . Our goal here is to introduce pattern recognition as the best way of utilizing available sensors, processors and domain knowledge to make automatic decisions.

II PATTERN RECOGNITION APPROACHES

2.1 Statistical Approach

Each pattern is represented in terms of d features and viewed as point in d-dimensional space and here my goal is to choose features which allows vector patterns belonging to different categories in order to occupy compact and disjoint regions in d-dimensional feature space. We can also determine the effectiveness of representation space by how well patterns from different classes can be separated [3] . In statistical region theoretical approach, decision boundaries are determined by probability distributions of pattern belonging to each class, which must either be specified or learned.

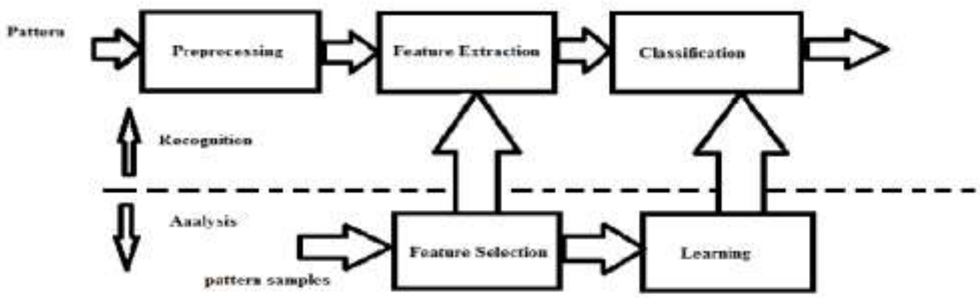

Fig. 1 Block Diagram of Statistical pattern Recognition System

2.2 Syntactic Approach

In various recognition problems these having complex patterns, it is more empirical to adopt hierarchial perspective where a pattern is viewed as being composed of simple sub-patterns which are themselves built from yet simpler sub-patterns. Large collection of complex patterns can be described by small no. of primitives and grammatical rules [2] . Grammar for each pattern class must be inferred from available samples set for training.

2.3 Structural Approach

Structural pattern recognition is intuitively appealing because, in its classification , this approach provides a description of how given pattern is constructed from primitives. This structural situation has been used inthose situations where pattern have a definite structure which can be captured in terms of a set of rules such as EKG waveforms, textured images and shape analysis of contours [4] . Another approach named as syntactic approach is implemented leads to many difficulties which primarily has to do with segmentation of noisy patterns. It may yield to a combinational explosion of possibilities to be investigated, demanding large training sets and very large computational efforts.

III BACK PROPAGATION ALGORITHM

Back-propagation is the general form ofWidrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and corresponding targeted vectors are used to give a training to a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities. Standard back-propagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term *back-propagation* refers to the manner in which the gradient is computed for nonlinear multilayer networks. There are a number of variations on the basic algorithm that are based on other standard optimization techniques, such as conjugate gradient and Newton methods. The simplest implementation of back-propagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly, the negative of the gradient. On iteration of this algorithm can be written as : $x_{k+1}= x_k- \alpha_k g_k$

where$x_k$is vector of current weights and biases,$g_k$is the current gradient, and α_k is the learning rate. The various back-propagation algorithms [1] are:

3.1 Batch Gradient Descent

Back-propagation is used to calculate derivatives of performance with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent: dX= Learning Rate* d(Performance)/dX

3.2 Gradient Descent with Momentum

Gradient descent with momentum, allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Gradient Back-propagation is used to calculate derivatives of performance with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent with momentum.
dX_n= momentum constant*dX_{n-1}+Learning Rate*d(Performance)/dX

3.3 Gradient Descent with Adaptive Learning Rate

With standard steepest descent, the learning rate is held constant throughout training. The performance of the algorithm is very sensitive to the proper setting of the learning rate. The performance of the steepest descent algorithm can be improved if we allow the learning rate to change during the training process. An adaptive learning rate attempts to keep the learning step size as large as possible while keeping learning stable. The learning rate is made responsive to the complexity of the local error surface. Back-propagation is used to calculate derivatives of performance with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent: dX = Learning Rate*d(Performance)/dX

3.4 Gradient Descent Momentum with Adaptive Learning Rate

It is described in the same way asGradient Descent with Adaptive Learning Rate, except that it has the momentum coefficient as an additional training parameter. Back-propagation is used to calculate derivatives of performance with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent with momentum,: dX_n= momemntum constant*dX_{n-1}+Learning Rate*momentum constant*d(Perf.)/dX

3.5 Resilient Back-propagation

The purpose of the resilient back-propagation training algorithm is to eliminate these harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative can determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update. Back-propagation is used to calculate derivatives of performance with respect to the weight and bias variables X. Each variable is adjusted according to the following: dX = deltaX*sign(Gradient)

3.6 Conjugate Gradient Descent Algorithm

In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions. In most of the conjugate gradient algorithms, the step size is adjusted in an iteration. A search is made along the conjugate gradient direction to determine the step size that minimizes the performance function along that line.

IV NEURAL NETWORK

Neural Network can be seen as massively parallel computing systems consisting an extreme large no. of processors which can be simple in nature with many interconnections. Neural Network models attempt to use some organizational principles as learning rate, generalization, fault tolerance, adaptivity, distributed representations, computations and so on in a network of weighted directed graphs in which the nodes defined by artificial neurons and directed edges are the connections between neurons output and neurons inputs. The most common family used in neural networks for pattern classification tasks is the Feed-Forward network[5] , which includes multilayer perceptron and Radial-basis function RBF networks. It provides a new suite of non-linear algorithm for feature extraction and classification algorithm can be mapped on neural network architecture for efficient and effective implementation.

Error Estimation

While designing neural network, these parameters must be decided.No. of neurons in hidden layers, Learning Rates, Momentum, Training Types, Epoch, Minimum Error

4.1 Mean Squared Normalized Error Performance Function (MSE)

Mean Squared Error is the average squared difference between outputs and targets.MSE is the second moment of the error, and thus incorporates both the variance with respect to target value. The MSE of an output value ŷ with respect to the target value y is defined as

$$MSE(\hat{y}) = E[(\hat{y}-y)^2]$$

4.2 Estimating Number of Components

ML criterion cannot be used to estimate the number of mixture components because maximized likelihood is a non-decreasing function of K, so thereby making it useless as a model selection criterion with a more "fully Bayesian flavor" to sample from the full posteriori distribution where K is included as an unknown. We strongly think that MCMC-based techniques are still far too computationally demanding to be useful in pattern recognition application. K means it will not be able to identify three components, due to substantial overlap of two components. K is selected using modified MDL criterion.

V CONCLUSION

Comparative view has been shown of all pattern recognition models and these models depict various domains in this perspective area of research. Different and combination of various other models can also be used. In case of noisy model patterns the choice of statistical model will be a good solution. And the experimental knowledge of structural model depends upon recognition of simple pattern primitives and their relationships presented by descriptive languages. As neural network and statistical pattern recognition models have different principles. To recognize unknown shapes or patterns the only best method is to use fuzzy methods, therefore to enhance system performance for complex applications it is beneficial to appendmore than one recognition model at various stages of recognizing models.

REFERENCES

[1]Ritika, Handwritten numeral pattern recognition in neural network, Theses Report"BMIET 2016, ICRISEM-2016.

[2] A.L. Knoll, "Experiments with 'Characteristic loci' for recognition of handprinted characters," IEEE Trans. Computers, vol 18, pp. 366-372, 1969

[3]P. Mermelstein and M. Eden, "Experiments on computer recognition of connected handwritten words, "Inform Contr. ,vol 7.pp. 255-270, 1964

[4] B. Gatos, I. Pratikakis and S.J. Perantonis, "Hybrid Off-Line Cursive Handwritten Word Recognition ", 18th International Conference on Pattern Recognition (ICPR2006),pp. 999, Hong Kong, August 2006.

[5] T. Ahmad, A. Jameel, Dr. Balal Ahmad, "Pattern Recognition using statistical and Neural Techniques", 2011 IEEE.