



Pattern Recognition for Diabetic Spectral Data Using Machine Learning Approaches

KEYWORDS

Machine Learning, Artificial Neural Network, Radial Basis Function Network, Multilayer Perception Neural Network, diabetic Spectral Data.

Dharuman C

SRM University, Ramapuram, Chennai-89.

Venkatesan P

National Institute for Research in Tuberculosis, Chennai-31

ABSTRACT A vector representation of spectra leads to high dimensional problems. Hence the Machine Learning approaches are required for the analysis of the high dimensional diabetic Spectral data. The Multilayer perception (MLP) Radial Basis Function Network Support Vector Machine and Logistic Regression Model were applied to the diabetic spectra data. The efficiency of the models are evaluated by sensitivity, specificity and accuracy. The results show that the Neural Network Models perform better than Logistic Regression and Support Vector Machine out performs the ANN Models.

1. INTRODUCTION

Machine learning is a new technology with a wide range of applications. It has become one of the key components of intelligent information systems, enabling compact generalizations, inferred from large databases of recorded information. It can be applied as knowledge in various practical ways such as being embedded in automatic processes like expert systems (Venkatesan and Anitha 2006, Venkatesan and Suresh 2009). Data mining is a commercially and scientifically important area of application where algorithms are used to detect relevant information and patterns in large databases (Mitchell 1997). Frequently used techniques include symbolic, inductive learning algorithms such as ID3, multiple-layered, feed-forward neural networks (Rumelhart et al., 1986), and evolution-based genetic algorithms (Goldberg 1989).

2. MACHINE LEARNING APPROACHES

Learning is an inherent characteristic of the human beings. By virtue of this, people, while executing similar tasks, acquire the ability to improve their performance. This paper provides an overview of the principle of learning that can be adhered to machines to improve their performance. Such learning is usually referred to as machine learning which can be broadly classified into three categories: i) Supervised learning ii) Unsupervised learning and iii) Reinforcement learning. The learner adapts its parameters based on the status of its actions. Among the supervised learning techniques, the most common are inductive and analogical learning. Hopfield networks have been used extensively in the area of global optimization and search (Hopfield 1982); Kohonen networks have been adopted in unsupervised learning and pattern recognition. The other important machine algorithms are support vector machines (Chu and Wang, 2003, Meyer et al., 2003), genetic algorithms (Goldberg 1989, Mitchell 1998) and fuzzy logic.

2.1 Artificial Neural Network

All connectionist algorithms have a strong learning component. Learning algorithms can be applied to adjust connection weights so that the network can predict or classify unknown examples correctly. Neural networks have been adopted in various engineering, business, military, biomedical and chemical domains (Simpson, 1990). These artificial neural networks have a number of successful applications in biology such as pattern recognition in DNA and proteins structure prediction analysis and clustering of gene expression data, modeling gene network (Bishop, 1995, Shawe-Taylor and Cristianini, 2004).

The FF net uses a supervised learning algorithm; besides the

input pattern, the neural net also needs to know to what category the pattern belongs. Learning proceeds as follows: a pattern is presented at the inputs. The pattern will be transformed in its passage through the layers of the network until it reaches the output layer. The units in the output layer all belong to a different category. The outputs of the network as they are now compared with the outputs as they ideally would have been if this pattern were correctly classified. The differences between the actual outputs and the idealized outputs are propagated back from the top layer to lower layers and connection weights modified to get the desired output. A typical neural network will be as given in Fig.1

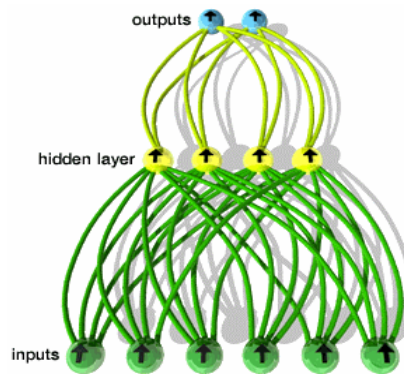


Fig.1 Basic model of a network.

Mathematically the functionality of a hidden neuron is described by

$$\hat{o} = \sum_{j=1}^n w_j x_j + b_j \quad (1)$$

where the weights w_j, b_j are symbolized with the arrows feeding into the neuron. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer. The output of this network is given by

$$\hat{y}(\theta) = g(\theta, x) = \sum_{i=1}^{nh} w_i^2 \sigma \left[\sum_{j=1}^n w_{ij}^1 x_j + b_{ij}^1 \right] + b^2 \quad (2)$$

where n is the number of inputs and nh is the number of neurons in the hidden layer. The variables w_j^1, b_j^1, w_i^2, b^2 are

the parameters of the network model that are represented collectively by the parameter vector θ . In general, the neural network model will be represented by the compact notation $g(\theta, x)$. In training the network, its parameters are adjusted incrementally until the training data satisfy the desired mapping as closely as possible; that is, until $\hat{y}(\theta)$ matches the desired output y . The nonlinear activation function in the neuron is usually chosen to be a smooth step function given by

$$\text{sigmoid}[x] = \frac{1}{1 + e^{-x}} \tag{3}$$

2.2 Back propagation algorithm

Back propagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. It was first described by Werbos in 1974, but it wasn't until 1986, through the work of Rumelhart, et al., (1986) that it gained recognition, and it led to a renaissance in the field of artificial neural network research. It is a supervised learning method, and is an implementation of the Delta rule. It can calculate the desired output for any given input. It is most useful for feed-forward networks.

2.3 Radial Basis Function Network

Radial basis function (RBF) networks have a static Gaussian function as the nonlinearity for the hidden layer processing elements. The Gaussian function responds only to a small region of the input space where the Gaussian is centered. The simulation starts with the training of an unsupervised layer. During the unsupervised learning, the widths of the Gaussians are computed based on the centers of their neighbors. The output of this layer is derived from the input data weighted by a Gaussian mixture. Once the unsupervised layer has completed its training, the supervised segment then sets the centers of Gaussian functions and determines the width of each Gaussian. Any supervised topology may be used for the classification of the weighted input. The advantage of the radial basis function network is that it finds the input to output map using local approximators.

Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer are given in Fig.2 and 3

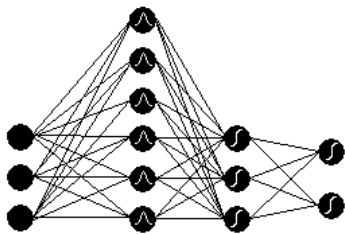


Fig. 2 Radial Basis Function Network

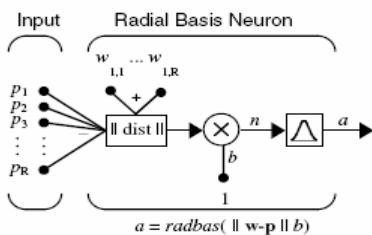


Fig.3 Neuron Model with R inputs

The output, $\varphi = \mathbb{R}^p \rightarrow \mathbb{R}$, of the network is given by

$$\varphi(x) = \sum_{i=1}^N a_i \rho(\|x - c_i\|) \tag{4}$$

where N is the number of neurons in the hidden layer, C_i is

the center vector for neuron i , and a_i are the weights of the linear output neuron. In the basic form all inputs are connected to each hidden neuron. The norm is typically taken to be the Euclidean distance and the basis function is taken to be Gaussian given by

$$\rho(\|x - c_i\|) = \exp(-\beta\|x - c_i\|^2) \tag{5}$$

The Gaussian basis functions are local in the sense that is

$$\lim_{\|x\| \rightarrow \infty} \rho(\|x - c_i\|) = 0 \tag{6}$$

i.e., changing parameters of one neuron has only a small effect for input values that are far away from the center of that neuron. The weights a_i , c_i and β are determined in a manner that optimizes the fit between φ and the data.

2.5 Supporting Vector Machine

Support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Formally, a SVM constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class, since in general the larger the margin the lower the generalization error of the classifier.

The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category or the other. A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Given some training data D , a set of n points of the form

$$D = \{(X_i, Y_i) \mid X_i \in \mathbb{R}^p, Y_i \in \{-1, 1\}\}_{i=1}^n \tag{7}$$

where the y_i is either 1 or -1, indicating the class to which the point X_i belongs. Each X_i is a p -dimensional real vector. We want to find the maximum-margin hyper plane that divides the points having $Y_i = 1$ from those having $Y_i = -1$.

where \cdot denotes the dot product and \mathbf{W} the normal vector to the hyper plane.

If the training data are linearly separable, we can select two hyper planes in a way that they separate the data and there are no points between them, and then try to maximize their distance. The region bounded by them is called "the margin". These hyper planes can be described by the equations

$$\mathbf{w} \cdot \mathbf{x} - b = 1 \tag{8}$$

$$\text{and } \mathbf{w} \cdot \mathbf{x} - b = -1. \tag{9}$$

where the parameter b determines the offset of the hyper plane from the origin along the normal vector \mathbf{w} . We find the distance between these two hyper planes is $\frac{2}{\|\mathbf{w}\|}$, so we want to minimize $\|\mathbf{w}\|$. As we also have to prevent data points from falling into the margin, we add the following constraint: for each i either

$$\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \quad \text{for } \mathbf{x}_i \text{ of the first class} \tag{10}$$

(or) $w \cdot x_i - b \leq -1$ for x_i of the second (11)

3. APPLICATION TO FTIR PATTERN ANALYSIS

FTIR Spectroscopy is a form of vibrational spectroscopy and the spectrum reflects both molecular structure and molecular environment (Zellar et al., 1989).. A molecule when exposed to the radiation produced by the thermal emission of a hot source absorbs only at frequencies corresponding to it's molecular mode of vibration in the region of the electromagnetic spectrum between visible and short waves. These changes in vibrational motion give rise to bands in the vibrational spectrum; each spectral band is characterized by its frequencies and amplitude.

4. Materials and Methods

The Data Base consists of 18 Diabetes and 11 other non Diabetes. The identification of them is usually carried out by noticing the characteristic stretching vibrations. Based on the intensity and location of peak, the change of chemical structure can be identified. Each spectrum consists of much absorption and those appearing on the finger print region (1600-900 cm^{-1}) are very characteristic of the molecule. The other signals in the region (1600-400 cm^{-1}) are mostly contributed by functional groups. The bending vibrations appearing in the region (900-400 cm^{-1}) are also useful for understanding the finer details of the molecules. The Overlay plot of the FTIR Spectra of Diabetes and non-Diabetes Data glucose region (1250 – 925 cm^{-1}) in Fig.4

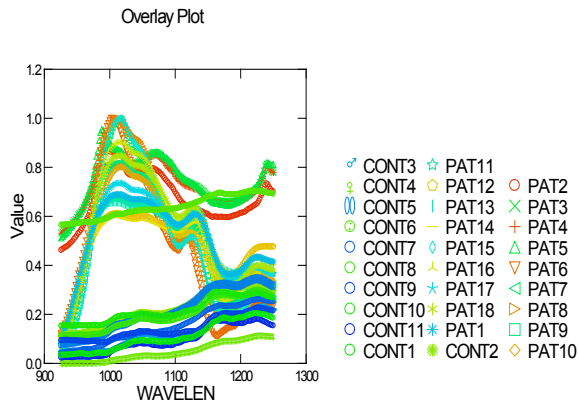


Fig 4 : Overlay plot of the FTIR Spectra of Diabetic and Non-Diabetic data (925-1250 cm^{-1})

Using PCA the features were extracted for using as input in the ANN. The three components estimate about 99.25% of the information as in Table 1. The data set is small and we have divided 75 percent for training and 25 percent for testing MLP and RBF networks. The binary logistic regression analysis was also fitted for comparison. All the models were fitted using SPSS 16.0 package. The efficiency of the models is evaluated by sensitivity, specificity and accuracy.

For MLP network architecture, a single hidden layer with sigmoid activation function, which is optional for the diclorotomone out, is chosen. A back propagation algorithm based on conjugate gradient optimization technique was used the model MLP. The RBF network considered for this application was a single hidden layer with Gaussian kernel and the activation function used is symmetric. The cross validation layer error correction method is used (Paola and Schowengerdt, 1994). The logistic regression model was fitted using the same input vectors as in the neural network.

5 Results

PCA was carried out to extract the feature vectors latent roots. The first three latent components of PCA account for more that 99%. The results are presented in Table 1 and 2

These components are used as inputs neural networks to

construct both ML and RBF, The FFNN architecture of diabetic and organic data consists of one hidden layer three nodes for MLP network and four nodes for RBF networks. The results are presented in Figure 5 and Figure 6 and weights in Tables 3 and 4. The numbers of nodes in the hidden layer are two organic data. The sigmoid activation function is used for hidden and output nodes. The input vectors are rescaled using standardized method and output units are rescaled using normalization method. The error function used is sum of squares.

Table 1 FTIR Diabetic Data - Factor analysis

Latent Roots	Variance Value	Total Variance (%)	Cumulative (%)
1	133.07	81.14	81.14
2	27.25	16.62	97.76
3	2.44	1.49	99.25

Table 2 Component loading for selected wave numbers

Wave numbers	Components		
	PC ₁	PC ₂	PC ₃
1250	0.878	-0.452	0.071
1226	0.891	-0.444	0.004
1224	0.891	-0.444	-0.002
1210	0.871	-0.487	-0.011
1144	0.956	-0.049	-0.281
1142	0.960	0.000	-0.272
1140	0.962	0.044	-0.264
1162	0.914	0.399	-0.002*
1160	0.914	0.399	0.003
1058	0.908	0.414	0.009
960	0.986	0.017	0.058
958	0.985	-0.031*	0.069
956	0.981	-0.078	0.086
926	0.737	-0.575	0.329

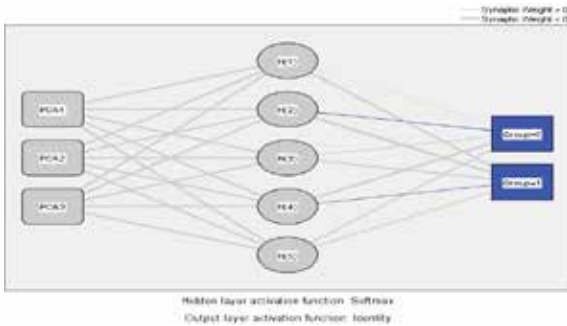
Hidden layer activation function: Sigmoid
Output layer activation function: Sigmoid



Fig.5 FFNN model for Diabetes data

Table 3 Parameter estimates for FFNN model (Diabetes data)

Predictor		Predicted				
		Hidden Layer 1			Output Layer	
		H(1:1)	H(1:2)	H(1:3)	[Group=0]	[Group=1]
Input Layer	(Bias)	3.585	-.024	-3.675		
	PCA1	3.644	.665	-2.195		
	PCA2	3.594	1.480	-2.233		
	PCA3	3.713	1.647	-3.091		
Hidden Layer 1	(Bias)				-1.052	.133
	H(1:1)				-2.621	2.388
	H(1:2)				-.662	.741
	H(1:3)				2.406	-2.632



Hidden layer activation function: Softmax

Output layer activation function: Identity

Fig.6 RBFN model for Diabetes data

Table 4 Parameter Estimates for RBFN Model (Diabetes data)

Predictor		Predicted					Output Layer	
		Hidden Layer ^a					[Group =0]	[Group =1]
		H(1)	H(2)	H(3)	H(4)	H(5)		
Input Layer	PCA1	1.998	.088	-.446	-.957	-.592		
	PCA2	2.000	.084	-.445	-.955	-.592		
	PCA3	2.003	.081	-.442	-.954	-.594		
Hidden Width	Unit	.091	.253	.213	.188	.075		
Hidden Layer	H(1)						3.001E-38	1.000
	H(2)						-.015	1.015
	H(3)						.266	.734
	H(4)						1.011	-.011
	H(5)						.682	.318

a. Displays the center vector for each hidden unit

Table 5 Comparative predictions of the Models

Database	Model	Sensitivity (%)	Specificity (%)	Correct Prediction (%)
Diabetes data	Logistic	81.8	77.8	80.3
	MLPNN	81.8	83.3	82.4
	RBFNN	90.9	88.9	90.1
	SVM	90.9	94.4	92.2

Both the network models gave higher sensitivity and specificity compared to logistic model. Among the network models the RBFNN gave higher prediction compared to MLPNN. SVM performs better than the neutral networks models

3.3 Discussion

The back propagation approach, ANN, have a number of advantages over traditional parametric approaches including the ability to model non-linear relationships, no specific assumption concerning the distributional characteristics and accommodation of variable interactions without a priori specification: ANN methodologies being with a fully connected with neural network composed of neurons with completely random connection weights that link the inputs to outputs: The back propagation algorithm modifies the connections weights in an interactive fashions to maximize the match between predicted and observed levels. The optimal network configuration was determined by comparing the performances of different 10-fold cross-validated network with 1 to 10 hidden neurons. The net work performance classifying bromide a Spectral data is similar to the above results. (Venkatesan et al.2011). The sum outperforms the neural networks. This was reported by several others. The was However more studies are needed using different FTIR Chemical Data base to confirm findings.

REFERENCE

1. Bishop, C M : Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995. | 2. Chu, F and Wang, L : Gene expression data analysis using support vector machines, Proceedings of the 2003 IEEE International Joint Conference on Neural Networks, Portland, USA, 2003, 2: 2268–2271. | 3. Goldberg, D E : Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA, 1989. | 4. Hopfield, J J : Neural network and physical systems with collective computational abilities, Proceedings of the National Academy of Science, 1982, USA, 79:2554-2558. | 5. Mitchell, M : Machine Learning, ISBN 0-07-042807-7, McGraw-Hill, 1997. | 6. Meyer, D., Leisch, F., and Hornik, K : The support vector machine under test, Neuro computing 55, 2003, 2: 169-186. | 7. Paola J D and Schowengerdt R A : Comparison of neural network to standard techniques for image classification and correlation. In Proc. Int'l Geosci. And Remote Sensing Symposium, volume III, pages 1404-1405, Caltech, Pasadena, CA, August 8-12, 1994. | 8. Shawe-Taylor, J and Cristianini N : Kernel Methods for Pattern Analysis and Discovery, Cambridge University Press, 2004. | 9. Simpson, P K : Artificial Neural Systems : Foundations, Paradigms, Applications, and Implementations. McGraw-Hill Book Company, New York, 1990. | 10. Venkatesan, P and Anitha, S : Application of a radial basis function neural network for diagnosis of diabetes mellitus, Current Science, 2006, 91:1195-1199. | 11. Venkatesan, P and Suresh, M L : Breast Cancer Survival prediction using ANN, International Journal of Computer Science and Net Work Security, 2009, 9: 169-1 | 12. Venkatesan, P, Dharuman, C and Gunasekaran, S : Multivariate Pattern Recognition Approach for Chemical FTIR Spectra. Recent Trends in Statistics and Computer Applications, Manonmaniam Sundarnar University, 2011, pp 281-286.