ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Classifiers for the Epilepsy Risk Level Classification from Electroencephalographic Signals.

R Harikumar, and P Sunil Kumar*.

Department of ECE, Bannari Amman Institute of Technology, Tamil Nadu, India.

ABSTRACT

In the field of machine learning analysis and statistical theory, classification is actually the problem of identification of a new observation to a particular set of categories. It is done on the basis of a training set of a particular data which contains its respective observations and whose category membership is known. In the context of machine learning, classification is generally considered as an instance of supervised learning. When a particular training set of correctly identified observations are available, then it is known as supervised learning. An algorithm that implements classification in a perfect concrete implementation form is known as a classifier. Sometimes classifiers can also refer to any mathematical functions which could be implemented by any classification algorithm where mapping of an input data to a particular category is done simultaneously. This paper reviews the different types of classifiers used for the perfect classification of epilepsy risk levels from Electroencephalography signals.

Keywords: Machine learning, Statistical theory, Epilepsy

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Classifiers are able to recognize differences in patterns based on automatic learning procedures. The application is attractive because it not only provides faster responses, but also it has the capability to automatically discover irregularities in patterns which are not seen or detected before [1]. Another important feature is that it enables the discovery of regularities in the training signal itself as a consequence of the actual learning process. A training procedure can be performed by three different categories. Supervised training requires one or more supervisors and it labels the data to be used in teaching the network. The system, knowing the correct answers, inputs an error signal, the moment the network produces an incorrect response. It continues to do so by feeding the difference in assessments back into the network until the error has been decreased to a predetermined minimum value. The error signal as it was teaches the network the correct response. Unsupervised or self-organized training, using unlabelled training sets do not require a supervisor. Internal clusters, compressing offered data into classification categories are formed when the moment data is presented to the network. The supervisor is also absent in self supervised training. The error signals, generated by the network, are fed back into the network itself until a correct response is produced. Such a type of methodology is being employed by the classifiers. This paper gives an insight about the classifiers which could be used for processing epileptic encephalographic signals.

Classifiers used for processing epileptic encephalographic signals

Some of the most important used classifiers for the classification of epilepsy risk levels from EEG signals are Non-linear Bayesian classifiers, Nearest Neighbour classifiers, Statistical classifiers, Machine Learning classifiers and Neural Network classifiers.

Non-linear Bayesian Classifiers

The non-linear Bayesian Classifiers included in this paper are Hidden Markov Model and Bayesian Quadratic Model.

Hidden Markov Model (HMM)

Basically it is a statistical markov model. The system which is modeled here is assumed to have a markov process and it has many unobserved or hidden states. In a HMM model, the state is not directly visible to the observer. The output which is usually dependent on the state is visible. Each and every state has a particular probability distribution function and it is taken over the possible output tokens. Therefore, the HMM generates the sequence of tokens and some information is provided with the sequences of states. HMM can be used widely in temporal pattern recognition such as speech, bio-signal processing and bio-informatics etc.

Bayesian Quadratic Model (BQM)

It is a standard approach employed for many supervised classification problems. The likelihood of each class is modeled as a Gaussian distribution and then it uses the posterior distribution for the estimation of the class for a given test point. Bayesian estimation for quadrant discriminant analysis has become very popular as it minimizes the cost of the expected misclassification.

Nearest neighbor classifiers

KNN Classifier:

It is an instance-based classifier. They always act on the assumption that classification of unknown instances can be easily done if the unknown is related to the known. Relating the unknown parameters to the known parameters can be done easily by the distance/similarity function. Since it is an instance based learning technique, it does not abstract any information from the training data during the learning phase. The generalization process would be postponed until it is absolutely mandatory or unavoidable at the time of classification. Since it does not hold any data in the learning phase, it is called as a lazy learner.

Mahalonobis distance

It is nothing but the measure of the distance between a particular point A and at distribution D. The idea behind this concept is multi-dimensional generalization and it measures how many standard deviations of A are away from the mean of D. Sometimes the Mahalonobis distance corresponds to the standard Euclidean distance under transformed space conditions. Mahalonobis distance is scale-invariant and it is also unitless and it generally takes into account the correlations of the particular data set.

Statistical Classifiers

Fischer's Linear Discriminant Analysis (FLDA)

Linear Discriminant Analysis (LDA) is the generalized process of Fischer's linear discriminant and it is used mostly in statistics, mathematics and machine learning algorithms. The main idea is to trace the linear combination of features that separates many different classes of events. It also attempts to combine the linear combination of other features as one dependent variable. Because of its special traits, it is closely related to the concepts like Analysis of Variance and Regression analysis. The independent variables for each observation should be for continuos quantities and only then LDA can work well. If the categorical independent variables are dealt, then the corresponding technique used is discriminant correspondence analysis.

Decision trees

It is basically a decision support tool. It mostly uses a tree shaped model or representation to explain its decisions which includes the possible event outcomes, utility and resource costs [5]. Decision trees are most common in signal processing, image processing, pattern recognition, operation research, decision analysis etc… Generally, the decision trees would consist of three types of nodes such as decision nodes, chance nodes and end nodes.

Naïve Bayes Classifier (NBC)

It belongs to the family of simple probabilistic classifiers and it is actually a conditional probability model. When Bayes theorem is applied with strong independent assumptions in between the features, then it is called as Naïve Bayes classifier in machine learning algorithm. Though this classifier is used in many arenas, its vital application is in the area of text categorization and it can be applied for the classification of epilepsy risk levels from EEG signals. With the correct application of the preprocessing techniques, it remains more competitive than the support vector machines and its application is of great use in the field of automatic medical diagnosis.

Machine Learning Classifiers

Support Vector Machine (SVM)

In machine learning, it is one of the supervised learning models that is used to analyze the data and recognize the patterns which is primarily used for the classification and regression analysis. It is a non-probabilistic binary linear classifier. The SVM training algorithm generally builds a model that frequently assigns new examples into one category or the other. SVM's can also perform a non-linear classification known as kernel-trick, where to a high-dimensional feature space, mapping of inputs is done implicitly.

Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) for Single –Hidden Layer Feed Forward Neural Network (SLFNS) are either additive neurons or Kernel based schemes. For additive neurons based SLFNS, one may randomly choose and fix the input weights (linking the input layer to the hidden layer) and the hidden neurons biases and analytically determine the output weight (linking the hidden layer to the output layer) of SLFNs. Input weights are the weights of the connections between input neurons and hidden neurons and output weights are weights of the connection between hidden neurons and output neurons. After the input weights and the hidden layer biases are chosen arbitrarily, SLFNs can be simply considered as a linear system and the output weights are analytically determined through simple generalized inverse operation of the hidden layer output matrices. ELM tends to have good generalization performance and can be implemented easily. Unlike other tuning or adjustment methods which may neither be suitable for non-differential activation functions nor prevent the troubling issues such as stopping criteria, learning rate, learning epochs, and local minima the ELM algorithm can avoid these difficulties very well.

Neural Network Classifiers

The EEG records can be interpreted and analyzed in many aspects. During the years it became clear that conventional (artificial intelligence oriented) approaches, based on obeying a set of rules, are not very suitable to perform such a task. The goal would be to develop a system, which takes the EEG record and clinical data as its inputs, and produces a judgment, whether the record is abnormal, and a clinical interpretation as its outputs [2]. Using raw EEG time series as an input for an Artificial Neural Network (ANN) is technically unfeasible. So, some form of processing and data reduction is required [3]. A logical choice would be to use the power contents of a few frequency bands as an input to ANN. However, power spectra may not be suitable to detect transient epileptic activity. Alternatively, some form of Gabor transformation must be used [4]. Another difficult problem to be addressed is the choice of the overall architecture. A collection of networks is trained for a specific well defined task. This would make the system flexible because new tasks could be added later on, and individual networks could be changed without affecting the architecture of the whole system.

Supervised Learning

The commonly used supervised learning techniques are Recurrent Neural Network (RNN), Wavelet Neural Network (WNN), Probabilistic Neural Network (PNN), Back Propagation Network (BPN), Multi Layer Perceptron (MLP) and Radial Basis Function (RBF).

Recurrent Neural Network

It is a type of Artificial Neural Network where connections between the units form a directed cycle. An internal state of the network is created thereby allowing the behavior to exhibit a dynamic and temporal nature. The arbitrary sequence of inputs can be easily processed using the internal memory of the RNN. This special trait makes them most applicable to tasks such as handwriting recognition where the best known results could be easily achieved.

Wavelet Neural Network

It combines the theory of wavelets and neural networks into one. The wavelet neural network generally comprises of feed-forward neural networks. It would usually have only one hidden layer and the activation functions would be drawn easily from the orthonormal wavelet family. The most important application of wavelet neural network is that of function estimation. If a series of observed values of a particular function is given, then the wavelet network can be used to train and learn the composition of that particular function and hence the expected value for a given input could be easily calculated.

Probabilistic Neural Network

It is also a feed-forward neural network, which is derived mainly from both the Bayesian network and a statistical algorithm called Kernel Fischer Discriminant analysis. In a PNN, each and every operation is organized into a particular multilayered feed forward network with four layers respectively such as input layer, pattern layer, summation layer and output layer.

Back Propagation Network

For training neural networks, this is a very commonly used method. It is always used in combination with an optimization technique such as gradient descent algorithm. The gradient of a particular loss function is calculated easily when considering all the weights in the network. To update the weights, the gradient is always fed to the optimization method thereby the loss function is always minimized. For each and every input value, back propagation algorithm always require a known and desired output value so that the calculation of the loss function gradient would be easier. In general, it is considered to be a supervised learning method whereas in auto encoders, it can act as unsupervised learning method.

Multilayer Perceptron

Multilayer Perceptron is feed forward neural networks trained with the standard back propagation algorithm. They are supervised networks and so they require a desired response to be trained. They learn how to transform the input data into a desired response, so they are widely used for pattern classification. Most neural network applications involve MLPs and very powerful pattern classifiers. With one or two hidden layers they can approximate virtually any input-output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems. They can efficiently use the information contained in the input data. The advantage of using this network resides in its simplicity and the fact that it is well suited for online implementation. The Levenberg-Marquardt (LM) algorithm is the standard training method for minimization of Mean Square Error (MSE) criteria, due to its rapid convergence properties and robustness. It provides a fast convergence, it is robust and simple to implement, and it is not necessary for the user to initialize any strange design parameters. It out performs simple gradient descent and other conjugate gradient methods in a wide variety of problems. This error back propagation algorithm is used to calculate the weight updates in each layer of the network.

Radial Basis Function

The Radial Basis Function neural network is widely used for function approximation, pattern classification and recognition due to its structural simplicity, universal approximators, and faster learning abilities due to locally tuned neurons. In its basic form, a RBF neural network involves three functionally distinct layers. The input layer is simply a set of sensory units. The second layer is a hidden layer of sufficient dimension which applies a non-linear transformation of the input space to a higher dimensional hidden layer. An RBF neural network is generally trained in two steps one after another. In the first step, the centers of hidden layer neurons are selected. Then the weights between the hidden and output layers are estimated. The centers of the hidden layer neurons of an RBF neural network are selected in different ways. Generally, these centers are selected by using some clustering algorithm like, k-means, fuzzy c-means, etc..

Unsupervised Learning

The commonly used unsupervised learning techniques are Self-Organizing Maps (SOM) and Adaptive Resonance Theory (ART).

Self-Organizing Maps

It is a type of ANN that is always trained using unsupervised learning. The main aim is to produce a low-dimensional data and also the input space of the training samples is represented in a discretized form called as map. They are unique from other ANN in the sense that in order to protect and preserve the topological properties of a particular input space, a neighborhood function is used. Similar to most of the ANN, SOM also operates in two modes, namely training and mapping. Building the map using input examples comes under training phase while classification of a new input vector automatically comes under the mapping phase.

Adaptive Resonance Theory

The basic ART concept is an unsupervised learning model. It basically comprises of four important modules namely, a comparison field, recognition field composed of many neurons, vigilance parameters and a reset module. The primary idea behind this ART model is that due to the interaction of 'top-down' observer expectations with 'bottom-up' sensory information, the recognition and object identification becomes much simpler. This model can be widely applicable under the case when new knowledge has to be acquired without the disruption of the existing knowledge.

CONCLUSION

Classifiers are effective in detecting and classifying patterns in all kinds of data sets. Recognizing these features has to lead a search for possible application of classifiers in medicine, but so far, only a few attempts have been made in applying these techniques to detect one specific type of activity in the EEG record. Classifiers can show excellent performance in detecting epileptic activities in EEG recordings. Classifiers are a powerful tool in pattern recognition problems also. Specifically, they are highly useful for automating diagnostic tasks carried out by experts as done in supervised classification tasks.

REFERENCES

[1] Celement C Pang, Upton ARM, Shinge G and Kamath MB. IEEE Transaction on Bio Medical Engineering 2003;50(4):521-26.
[2] Guoqiang Peter Zhang. IEEE Transactions on Systems Man Cybernetics - Part C: Applications and Reviews 2000;30(4):451-462.
[3] Haykin S. 'Neural networks a Comprehensive Foundation', Prentice- Hall Inc. 2nd Ed, New York, 1999.
[4] Mark van Gils, Rosen Falck A, White S, Piror P and Gade J. IEEE EMB Magazine 1997;16(6):56-63.
[5] Rasoul Safavian S. and David Landgrebe. IEEE Trans on SMC 1991;21(3):660-674.