A Survey On EEG Analysis Using Different Methodologies

B. Indira Priyadarshini, Dr. D. Krishna Reddy
Research Scholar, Professor
Osmania University
Chaitanya Bharathi Institute of Technology

***Abstract* – In this article we describe the analysis of EEG signal using different methodologies. How the different methodologies are able to achieve to extract the EEG signal and classify the signal. The survey defines the percentage of achievement of classification and extraction of signal.**

I. INTRODUCTION

The human brain is obviously a complex system and exhibits rich spatiotemporal dynamics. Among the noninvasive techniques for probing human brain dynamics, electroencephalography (EEG) provides a direct measure of cortical activity with millisecond temporal resolution. EEG is a record of the electrical potentials generated by the cerebral cortex nerve cells. There are two different types of EEG depending on where the signal is taken in the head: scalp or intracranial. For scalp EEG, the focus of this research, small metal discs, also known as electrodes, are placed on the scalp with good mechanical and electrical contact. Intracranial EEG is obtained by special electrodes implanted in the brain during a surgery. In order to provide an accurate detection of the voltage of the brain neuron current, the electrodes are of low impedance (<5 k_). The changes in the voltage difference between electrodes are sensed and amplified before being transmitted to a computer program to display the tracing of the voltage potential recordings. The recorded EEG provides a continuous graphic exhibition of the spatial distribution of the changing voltage fields over time. Epileptic seizure is an abnormality in EEG recordings and is characterized by brief and episodic neuronal synchronous discharges with dramatically increased amplitude. This anomalous synchrony may occur in the brain locally (partial seizures), which is seen only in a few channels of the EEG signal, or involving the whole brain (generalized seizures), which is seen in every channel of the EEG signal.

EEG signals involve a great deal of information about the function of the brain. But classification and evaluation of these signals are limited. Since there is no definite criterion evaluated by the experts, visual analysis of EEG signals in time domain may be insufficient. Routine clinical diagnosis needs to analysis of EEG signals. Therefore, some automation and computer techniques have been used for this aim. Since the early days of automatic EEG processing, representations based on a Fourier transform have been most commonly applied.

This approach is based on earlier observations that the EEG spectrum contains some characteristic waveforms that fall primarily within four frequency bands—delta (<4 Hz), theta (4—8 Hz), alpha (8—13 Hz) and beta (13—30 Hz). Such methods have proved beneficial for various EEG characterizations, but fast Fourier transform (FFT), suffer from large noise sensitivity. Parametric power spectrum estimation methods such as autoregressive (AR), reduces the spectral loss problems and gives better frequency resolution. But, since the EEG signals are non-stationary, the parametric methods are not suitable for frequency decomposition of these signals.

A powerful method was proposed in the late 1980s to perform time-scale analysis of signals: the wavelet transforms (WT). This method provides a unified framework for different techniques that have been developed for various applications [2—18]. Since the WT is appropriate for analysis of non-stationary signals and this represents a major advantage over spectral analysis, it is well suited to locating transient events, which may occur during epileptic seizures. Wavelet's feature extraction and representation properties can be used to analyze various transient events in biological signals. Adeli et al. [2] gave an overview of the discrete wavelet transform (DWT) developed for recognizing and quantifying spikes, sharp waves and spike-waves. They used wavelet transform to analyze and characterize epileptiform discharges in the form of 3-Hz spike and wave complex in patients with absence seizure. Through wavelet decomposition of the EEG records, transient features are accurately captured and localized in both time and frequency context. The capability of this mathematical microscope to analyze different scales of neural rhythms is shown to be a powerful tool for investigating small-scale oscillations of the brain signals. A better understanding of the dynamics of the human brain through EEG analysis can be obtained through further analysis of such EEG records. Numerous other techniques from the theory of signal analysis have been used to obtain representations and extract the features of interest for classification purposes. Neural networks and statistical pattern recognition methods have been applied to EEG analysis. Neural network detection systems have been proposed by a number of researchers Pradhan et al. used the raw EEG as an input to a neural network while Weng and Khorasani used the features proposed by Got based on these models are classified with a multilayer, feed

II. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are computing systems made up of large number of simple, highly interconnected processing elements (called nodes or artificial neurons) that abstractly emulate the structure and operation of the biological nervous

system. Learning in ANNs is accomplished through special training algorithms developed based on learning rules presumed to mimic the learning mechanisms of biological systems. There are many different types and architectures of neural networks varying fundamentally in the way they learn, the details of which are well documented in the literature. In this paper, neural network relevant to the application being considered (i.e., classification of EEG data) will be employed for designing classifiers, namely the MLPNN. The architecture of MLPNN may contain two or more layers. A simple two-layer ANN consists only of an input layer containing the input variables to the problem and output layer containing the solution of the problem. This type of networks is a satisfactory approximate or for linear problems. However, for approximating nonlinear systems, additional intermediate (hidden) processing layers are employed to handle the problem's nonlinearity and complexity. Although it depends on complexity of the function or the process being modeled, one hidden layer may be sufficient to map an arbitrary function to any degree of accuracy. Hence, three-layer architecture ANNs were adopted for the present study.

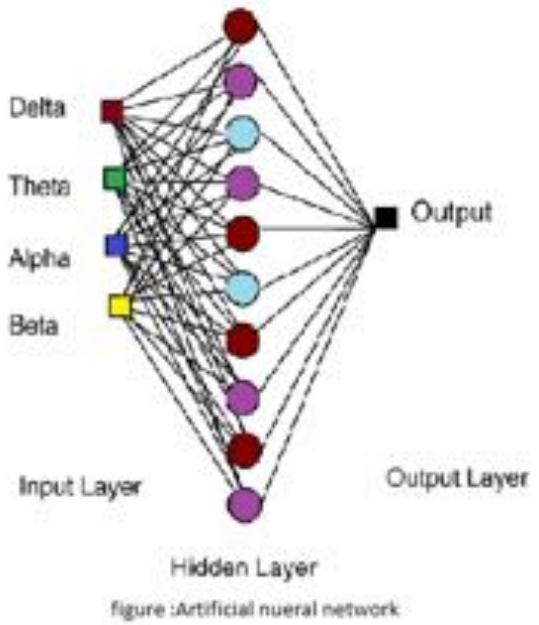

figure :Artificial nueral network

94 A. Subasi, E. Erc¸elebi structure of a fully connected three-layer network. The determination of appropriate number of hidden layers is one of the most critical tasks in neural network design. Unlike the input and output layers, one starts with no prior knowledge as to the number of hidden layers. A network with too few hidden nodes would be incapable of differentiating between complex patterns leading to only a linear estimate of the actual trend. In contrast, if the network has too many hidden nodes it will follow the noise in the data due to over-parameterization leading to poor generalization for untrained data.

With increasing number of hidden layers, training becomes excessively time-consuming. The most popular approach to finding the optimal number of hidden layers is by trial and error . In the present study, MLPNN consisted of one input layer, one hidden layer with 21 nodes and one output layer. Training algorithms are an integral part of ANN model development. An appropriate topology may still fail to give a better model, unless trained by a suitable training algorithm. A good training algorithm will shorten the training time, while achieving a better accuracy. Therefore, training process is an important characteristic of the ANNs, whereby representative examples of the knowledge are iteratively presented to the network, so that it can integrate this knowledge within its structure. There are a number of training algorithms used to train a MLPNN and a frequently used one is called the backpropagation training algorithm . The backpropagation algorithm, which is based on searching an error surface using gradient descent for points with minimum error, is relatively easy to implement. However, back propagation has some problems for many applications. The algorithm is not guaranteed to find the global minimum of the error function since gradient descent may get stuck in local minima, where it may remain indefinitely. In addition to this, long training sessions are often required in order to find an acceptable weight solution because of the well-known difficulties inherent in gradient descent optimization. Therefore, a lot of variations to improve the convergence of the backpropagation were proposed. Optimization methods such as second-order methods (conjugate gradient, quasi-Newton, Levenberg—Marquardt (L—M)) have also been used for ANN training in recent years. The Levenberg—Marquardt algorithm combines the best features of the Gauss—Newton technique and the steepest-descent algorithm, but avoids many of their limitations. In particular, it generally does not suffer from the problem of slow convergence.

III. MULTI-LAYER PERCEPTRON NEURAL NETWORK

Neural Network trained by a standard back propagation algorithm was used in our research. The number of neurons in the input layer varied according to the length of the input features vectors. Many tests were done to find the best configuration for the neural network in terms of: number of neurons in the hidden layer and the maximum number of iterations (epochs) in the learning process. For each features set, the configuration that produced optimal weights (which lead to maximum correct classification rate in the testing) for I/O mapping was used which were: Number of neurons in the hidden layer=100.

Maximum number of iterations (epochs) in the learning process=1000. The activation function used was the sigmoid function, the learning rate was 0.1 and the training stopped when either the maximum number of epochs reached 1000 or the mean

square error reached to a small value such as 0.001.

A.Support vector machine

The SVM classifier in this paper was based on LIBSVM implementation from [17]. Many tests were done to find the optimal parameters for SVM in terms of: type of the kernel, the Coefficient in kernel function, Degree in kernel function. Parameters that lead to maximum correct classification rate in the testing for I/O mapping were used which were:
Polynomial kernel was used.
Degree in kernel function=3.
Coefficient in kernel function=0.
Data were analyzed using MATLAB 2013 and a computer (Intel Core i7 CPU 2.20 GHz, 8 GB DDR RAM, Windows 7).
Total classification accuracies for classifying different combination of three mental tasks using the three feature extraction techniques and two classifiers as shown in Table 2 and Table 3 shows the effect of increasing the frequency band from [1 45] to [1 100] on it.
Table: Classification accuracies of different four mental tasks, frequency band [1 45].

	B,M,L,r	B,m,l,c	B,m,r,c	B,l,r,c	M,l,r,c
Svm fft	77.5	57.5	77.5	65	60
Svmwav	90	77.5	87.5	80	77.5
Svm pca	70	65	75	75	57.5
Nn fft	59	46.5	65.75	51.5	56.5
Nn wav	92.5	80	85.25	81.25	82.5

IV. INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent component analysis (ICA) is a method for the blind source separation problem. The use of ICA for blind source separation of EEG data is based on a plausible assumption that EEG data recorded at multiple scalp sensors are linear sums of temporally independent components arising from spatially fixed, distinct or overlapping brain or extra-brain networks [9]. The goal of ICA is to recover statistically independent sources given only sensor observations that are unknown linear mixtures of the unobserved independent source signals. In contrast to correlation-based transformations, ICA reduces the statistical dependencies of the signals, attempting to make the signals as independent as possible which makes ICA capable of isolating artifactual components from EEG data since they are usually independent of each other. Let x(t) represent n-dimensional vectors which correspond to the n continuous time series from the n EEG channels. Then xi(t) corresponds to the continuous electrode readings from the ith EEG channel. Because various underlying sources are summed via volume conduction to give rise to the scalp EEG, each of the xi(t) is assumed to be an instantaneous linear mixture of n unknown components or sources si(t), via the unknown mixing matrix A.
x(t) = A s(t)
ICA uses the EEG measurement x(t) and nothing else to generate an unmixing matrix W that approximates A-1, to recover a version, ŝ(t), of the original sources s(t), ŝ(t) = W x(t) (2)
Sensor space projections, which indicate the effect of a given component, in isolation, on all sensors are given by the estimated mixing matrix, Â = W-1 (3) Where each column of W-1 represents a spatial map describing the relative projection weights of the corresponding recovered source at each of the EEG channels: the first column of W-1 is the spatial map of the first independent component and so on. The columns of W-1 matrix are called the scalp topography (the scalp map) of the components which can be visualized in Fig. 3. These scalp maps provide evidence of the components' physiological origin (e.g. eye activity should project mainly to far frontal sites).
The use of ICA for removing artifacts from EEG signals is not new, However, in most previous work, it was used as an offline preprocessing method mainly for two reasons; the first one is related to the manual selection of artifactual components by visual inspection which is impractical, time consuming and the second is that ICA sufficient amount of data to be able to separate the independent sources. Therefore our goal was not only avoiding artifact rejection and minimizing data loss but also we wanted to correct EEG data automatically, as a very important step for an application such as BCI that require online and real time operation. Our ICA-based preprocessing technique consists of three steps:
(1) ICA decomposition. This is an offline training phase to obtain the unmixing matrix W, The idea is to train ICA on a whole 10-second trial and use the unmixing matrix W as a spatial filter to separate seven independent components from EEG data (the number of independent components is equal to the number of channels. We used Infomax (ICA), implemented with Mat lab 6p5 in EEGLAB interface [3], the 10-second trial was chosen arbitrarily, we chose the fifth trial of our subject's letter task shown in Fig. 2 (we note that this trial is not one of our six trials chosen for classification).
(2)Eye blink components identification. The preprocessing phase actually begins by applying the spatial filter W obtained from the training phase to the EEG data segments to separate the independent components or brain sources. Independent sources related to eye blinks must then be identified. The identification process is based on the components' scalp topographies. The scalp topography of each component provide evidence of its physiological origin. An eye blink component's scalp map has a strong far-frontal projection . A simple rule for eye blink components identification can be developed based on this fact and then these components must be removed from the components matrix ŝj(t) in order to reconstruct a clean EEG segment. Denoting the jth column of W-1 by W-1 j it represents the intensity distribution at each electrode (i.e. the scalp map) of the corresponding component.

EEG SIGNAL PREPROCESSING AND SEGMENTATION:

Information content of EEG signals is essential for detection of many problems of the brain and in connection with analysis of magnetic resonance images it forms one of the most complex diagnostic tools. To extract the most important properties of EEG observations it is necessary to use efficient mathematical tools [9], [10] to enable reliable and fast enough processing of very extensive data sets in most cases. Digital filters can be used in the initial stage of EEG data processing to remove power frequency from the observed signal and to reduce its undesirable frequency components. Presents a sample of a selected EEG channel comparing results of its segmentation by an expert and by a selected Bayesian method [2] detecting changes of its mean value and variation. This approach has been used in this case for a selected channel only even though further channels must be taken into account in the real case as well.

WAVELET ANALYSIS AND SIGNAL FEATURE EXTRACTION:

Wavelet transform forms a general mathematical tool for signal processing with many applications in EEG data analysis as well. Its basic use includes time-scale signal analysis, signal decomposition and signal compression. Abnormalities in EEG data during serious neurological diseases such as epilepsy are too subtle to be detected using conventional techniques that usually transform mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem. The techniques that have been applied to address this problem include the analysis of EEG signals for the detection of epileptic seizures using the autocorrelation function, time domain features, frequency domain features, time frequency analysis, nonlinear time series analysis, and wavelet transform. However, the results of various studies have demonstrated that the wavelet transform is the most promising method for extracting features from EEG signals. As such, the wavelet transform was used to extract features from EEG signals.

The wavelet transform, as a linear time-frequency transform, represents an efficient analytical tool in signal processing, pattern recognition and classification, and is suitable for analysis of transient and non-stationary phenomena as well as noise reduction. As a class of functions, it has the ability to localize information in both time and frequency . Therefore, the wavelet transform has been utilized widely in biomedical signal processing. In discrete wavelet analysis, a multi-resolution description is used to decompose a given signal into increasingly finer detail based on two sets of basic functions, the wavelets and the scaling functions, as follows:

$$x(t) = \sum_{k} 2^{j0/2}\, aj0(k)\varphi(2^j\, t - k) + \sum_{j=j0}^{\infty} \sum_{k} 2^{\frac{j}{2}} dj(t)\psi(2^j\, t - k)$$

where functions and are the basic scaling and mother wavelet, respectively. In the above expansion, the first summation represents an approximation of based on the scale index of , while the second term adds more detail using larger j (finer scales). The coefficients in this wavelet expansion are called the discrete wavelet transform (DWT) of the signal x(t).when wavelets are orthogonal these coefficients are

$$\int_{-\infty}^{\infty} 2^{\frac{j0}{2}} x(t)\varphi(2^j\, t - k)dt = aj(k)$$

$$\int_{-\infty}^{\infty} 2^{\frac{j}{2}} x(t)\psi(2^j\, t - k)dt = dj(k)$$

where a_j (k) and b_j (k) are the wavelet approximation and detail coefficients. The wavelet packet (WP) transform is a generalization of the DWT in which decomposition is undertaken in both directions (lower and higher frequencies). This general decomposition offers a greater range of possibilities for signal analysis than the discrete wavelet decomposition. In the WP tree, each node is recognized by the decomposition level (scale) l with respect to the WP tree root and the frequency band f. The ability of the wavelet transform in adaptive time-scale representation and decomposition of a signal into different frequency sub-bands presents an efficient signal analysis method without introducing a calculation burden [36]. Based on wavelet coefficients obtained after the wavelet transform, the signal can be reconstructed in each of the previously derived sub-bands and its time-domain features in different sub-bands can be studied separately.

Feature space reduction

After an appropriate signal analysis (e.g. wavelet transform used in this research), as well as feature extraction, the feature vector is derived. Its dimension should be reduced since the dimension is often too large and the design of classifiers for a large dimension suffers from various difficulties. Those are mostly numerical problems that involve operation with high-order matrices. At the same time, a classifier in -dimensional space is very difficult to analyze and almost impossible to imagine. Thus, it is helpful to define a matrix whose dimension is and in which the number of columns is smaller than the number of rows , such that the initial vector , following linear transformation , is projected onto the vector whose dimension is significantly smaller (e.g. 2 or 3 when it is possible to visualize classifiers in two- or three-dimensional space). Obviously, such transformation results in a loss of some information contained in the original vector but the classification procedure is simplified. The selection of the matrix is a trade-off between the desired level of simplicity of the classification procedure and the inevitable loss of information due to dimension reduction

Learning Algorithm for Structure Identification.

In our fuzzy neural networks, for every online incoming training pattern, we first use the novel cluster algorithm to identify the structure, and next apply the back propagation algorithm to optimize the parameters. In our learning method, only the training data is need. The input /output-term nodes and rule nodes are created dynamically as learning proceeds upon receiving on-line incoming training data. During the learning process, novel input-term and output-term nodes and rule nodes will be added. The

main idea of our clustering algorithm is for every input data, we first find the winner clusters in the input and output space respectively. Next, as in the fuzzy ARTMAP, we check that if the winner cluster in the input space is connected to the winner cluster in the output space. If so, we assume that the winner cluster in the output space is the correct prediction of the winner cluster in the input space, which is analogous to the fact the fuzzy ARTb category activated by the input is the correct prediction of the fuzzy ARTa categories activated by an input in the fuzzy ARTMAP.

V. CONCLUSION

In this paper we have presented the analysis of the EEG signal with different techniques and also we observed the behavior of the EEG signal.

REFERENCES

1. Blanchard G and Blankertz B (2004) 'BCI competition 2003-Data set IIa: Spatial patterns of self-controlled brain rhythm modulations' IEEE Transactions on Biomedical Engineering, Vol. 51, pp. 1062-1066.
2. Blume W, Lüders H, Mizrahi E, Tassinari C, van Emde Boas W, Engel J (2001) 'Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology', Epilepsia, Vol. 42, no. 9, pp. 1212–1218.
3. Brown, M., Grundy, W., Lin, D., Cristianni, W., Sugnet, C., Furey, T., Ares, M. and Haussler, D. (2000) 'Analysis of gene expression data by using support vector machines', Proc. Natl. Acad. Sci. Vol. 97 (1), pp.262-267.
4. Brunelli, R. (2009) Template Matching Techniques in Computer Vision: Theory and Practice, Wiley, New York.
5. Burke D P, Kelly S P, Chazal P, Reilly R B and Finucane C (2005) 'A parametric feature extraction and classification strategy for brain-computer interfacing' IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 13, pp. 12-17.
6. Burges, C. (1998) 'A tutorial on support vector machines for pattern recognition' Data Mining and Knowledge Discovery, Vol. 2. pp. 121-167.
7. Buzsaki, G. (2006) Rhythms of the brain, Oxford, Oxford University Press.
8. Caesarendra, W., Widodo, A. and Yang, B.S. (2010) 'Application of relevance vector machine and logistic regression for machine degradation assessment', Mechanical Systems and Signal Processing, Vol. 24, pp. 1161-1171.
9. Carlson, N. R. (2002a) Foundations of physiological psychology, 5th ed., Boston, Mass. London : Allyn and Bacon.
10. Cawley, G. C. and Talbot, N. L.C. (2008) 'Efficient approximate leave-one-out cross-validation for kernel logistic regression', Mach Learn, Vol. 71, pp. 243-264.
11. Carlson, N. R. (2002b) 'Structure and Functions of the Nervous System', Foundations of physiological psychology, Vol. 5th ed. Issue 3. Boston, Mass. London: Allyn and Bacon.
12. Chandaka, S., Chatterjee, A. and Munshi, S. (2009a) 'Cross-correlation aided support vector machine classifier for classification of EEG signals', Expert System with Applications, Vol. 36, pp. 1329-1336.
13. Chandaka, S., Chatterjee, A. and Munshi, S. (2009b) 'Support vector machines employing cross-correlation for emotional speech recognition', Measurement, Vol. 42, pp. 611-618.
14. Chiappa, S. and Millán, J.R. (2005) Data Set V<mental imagery, multi-class> [online]. Viewed 25 June 2009, http://ida.first.fraunhofer.de/projects/bci/competition_iii/desc_V.html
15. Cochran, W. G. (1977) Sampling Techniques, Wiley, New York. Collura, T. F. (1993) 'History and evolution of electroencephalographic instruments and techniques', Clinical Neurophysiology, Vol. 10, Iss. 4, pp. 476-504.
16. Congedo, M., Lotte, F. and Lecuyer, A. (2006) 'Classification of movement intention by spatially filtered electromagnetic inverse solutions', Physics inMedicine and Biology, Vol. 51, no. 8, pp.1971-1989.
17.De Veaux, R. D., Velleman, P.F. and Bock, D.E. (2008) Intro Stats (3rd edition), Pearson Addison Wesley, Boston.
18. Duda, R.O., Hart, P.E. and Stork, D.G. (2001) Pattern Classification, 2nd edn. John Wiley & Sons, New York.
19. Dutta S, Chatterjee A and Munshi S (2009) 'An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification', Expert systems,Vol. 26, pp. 202-217
20. Dutta, S., Chatterjee, A. and Munshi, S. (2010) 'Correlation techniques and least square support vector machine combine for frequency domain based ECG beat classification', Medical Engineering and Physics, Vol. 32, no. 10, pp. 1161-1169. EEG time series, 2005, [Online], Viewed 30 September 2008,
http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.htm