2016 2nd International Conference on Next Generation Computing Technologies (NGCT-2016)
Dehradun, India, 14-16 October 2016

Computer Aided Technique for Epilepsy Classification Using Cross wavelet Transform and RBF-Kernel Based Support Vector Machine

Sourya Sengupta, Debarshi Chanda, Anirjit Mitra
Jadavpur University, Kolkata, India
souryasengupta11@gmail.com

Saibal Dutta
Heritage Institute of Technology, Kolkata, India
saibal.dutta@heritageit.edu

Abstract**—** Fast and automatic identification and analysis of different bio-medical signals is of growing importance nowadays. This necessitates the application of different computer aided diagnosis methods to interpret, distinguish and analyze various signals and images. In this paper, we have proposed a novel method to identify the Epilepsy from EEG signals. RBF Kernel based Support Vector Machine (SVM) is employed for automatic classification of normal (with closed eyes) and epilepsy patients from their Electroencephalography or EEG signals. Six features are extracted from EEG signals using cross-wavelet transform. Cross-wavelet Transform has not been used before for EEG signal classification .These features are used to train SVM performing binary classification. The average accuracy of SVM based binary classifier is obtained as high as 84.90% in 10-fold cross-validation.

Keywords—EEG signal, Cross-wavelet, Cross-wavelet Spectrum, Support Vector Machine.

I. Introduction

EEG or Electroencephalography is an aperiodic time vs amplitude plot in which the information regarding the activity of the cerebral cortex nerve is obtained as a signal [1] . It is nothing but a monitoring method of brain signal. EEG has wide applications in diagnosis of different diseases. Epilepsy diagnosis is one of these applications . Although the actual cause of epileptic seizure is not an unique one but nevertheless it can be said that sudden and random seizure discharge of various brain neurons that temporarily hampers functions of the brain may lead to it. Globally there are 2.4 million new cases of epilepsy each year. Generally chance of occurrence of epilepsy is quite unpredictable, so the neuroscientists has considered the EEG signals as the most useful and easy way of studying brain's electrical responses. Epilepsy monitoring is done to distinguish epileptic seizures from different types of disorders and seizures, hence to classify the seizure types and find out whether an individual is epileptic or not. It may also be used to reach some other useful conclusions too. Earlier inspection and analysis of different EEG behaviors were purely based on human visualization. But it has been found that human observation is quite often very much error-prone and incapable for minute and close observations. So this evolves the need of computer aided automated tools in bio-medical signal processing [2]. Here we have used Crosswavelet Transform for feature extraction from EEG signals . Crosswavelet Transform has been used before for different biomedical applications but it has not been implemented for analysis of EEG signals. This method has some definite advantages over other methods as this method preserves temporal locality and it reveals localized similarities in time and scale of a signal.[3]

EEG signals can be interpreted by spectral analysis. Nowadays Artificial Neural Network is becoming popular for its superior performance than the spectral analysis for analysis of EEG signals [4]. Artificial Neural Networks can make decisions regarding different classes very effectively. Neural Networks work effectively in the field of different biological and medical applications, but now Support Vector Machine (SVM) is becoming popular in the various fields of signal processing, machine learning and pattern recognition. In many times SVM achieves higher accuracy for a particular problem than other classifiers.

A brief of the topics on which the different sections are concerned is given as follows. In section II information about EEG datasets used in our work is given. Section III depicts the method and the algorithm which we used in our work. Section IV gives an idea of cross-wavelet transform of two time domain signals. Section V presents the extraction of various features from the cross-wavelet transform. Section VI deals with the idea of support vector machine and how it works. Performance of the classifier scheme is shown in section VII. Section VIII describes the results obtained in this work and future research scope using EEG signal.

II. EEG Data Collection

We have used a publicly available EEG time series database [5]. All the signals which have been used are taken from a pre amplifier system having 128-channel. The data was digitalized with 12 bit resolution with a sampling rate of 173.61 samples per second. The database contains 5 sets of EEG signals which have been named as Set-A, Set-B, Set-C, Set-D and Set-E. Each of them has different significance [6]. Each of the dataset has EEG signals. All the signals have been taken from 100 single channel. Set A and set B of the total dataset have been

978-1-5090-3256-3/16/$31.00 ©2016 IEEE

taken from surface EEG recordings. Set A was for normal healthy patient with eyes open and set B was for normal healthy patients with eyes closed. For set C and D Signals were measured in seizure-free intervals from five patients. In the set C the signals were measured from the hippocampus formation of the opposite hemisphere of the brain .For set D it was in the epileptogenic zone. The last of the total dataset, set E has signals for seizure activity and these signals exhibit ictal activity.

Several works have been already reported regarding the classification of set A and set E signals .None of the previous works have used cross-wavelet transform for analysis. In our work we have tried a new technique for detecting epilepsy using cross wavelet transform and RBF-Kernel based SVM classifier . We have tested the result upon set B and set E. Different type of EEG signals are shown in Fig 1.X axis is time(in Seconds) and Y axis is amplitude(in V)

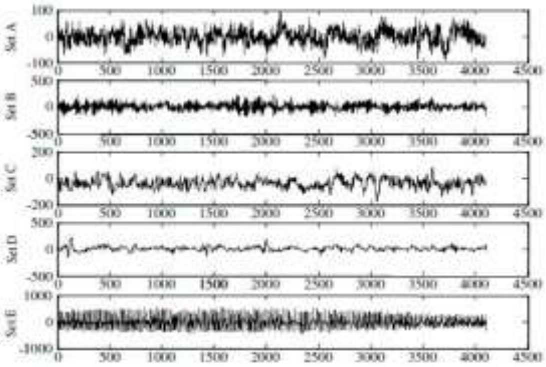

Fig 1: Different EEG signals

III. Proposed Methodology

In this work, we have used a fresh idea of feature extraction and classification. The flowchart of the procedure and steps followed is shown in Fig 2.

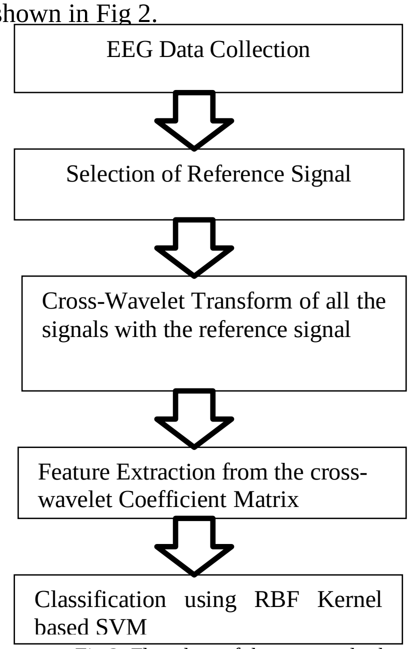

Fig 2: Flowchart of the proposed scheme

IV. Cross-Wavelet Transform

Wavelet analysis is an important tool nowadays for analyzing different signals [7]. It gives the measure of similarity between two different signals in three dimensions. It is basically the extended version of wavelet analysis which we usually see. The cross wavelet transform of two time domain signals $x_1(t)$ and $x_2(t)$ are given by:

$$W^{x1x2}(s,t) = \frac{1}{C_\varphi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W^{x1}(a,b)W^{x2}\left(\frac{a}{s},\frac{b-t}{s}\right).\frac{da.db}{a^2} \qquad (1)$$

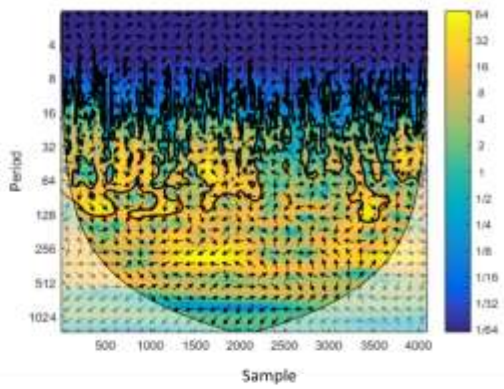

Fig 3: Cross-wavelet Spectrum of two signals from the dataset

Here $W^{x1}(s,t)$ is the wavelet transform of $x_1(t)$ and $W^{x2}(s,t)$ is the wavelet transform of $x_2(t)$. Both the transform are with respect to same mother wavelet, s and t are the default parameters commonly termed as dilation and translation parameters. In our work, Morlet mother wavelet is used for both the signals. However, other different mother wavelet can also be used. Here C_φ is a constant

We used a publicly available MATLAB package to compute cross-wavelet transform of the signals [8] .We have taken the first signal of the set B dataset as a reference signal and crossed the wavelet transform of this signal with the wavelet transform with all other signals. In fig 3, the sample cross-wavelet spectrum is shown.

V. Feature Extraction

Extraction of features is a very critical and determining step of the different classification projects. Here we have chosen few features among a variety of features available according to our needs. Accuracy of a particular problem largely depends on the selection of features. Feature extraction has been carried out using the feature stated in the literature [9,10]. After applying cross wavelet transform we a taken a six feature vector matrix (A_1-A_6) from the resulting cross-wavelet transform matrices.. All the features are described below:

A_1: $\frac{\sum_{i=1}^{s}\sum_{i=1}^{t} abs(X(i,j))}{Max}$

A_2: Median Value of X

A_3: Maximum value of i for which $X(i,1)$ is maximum

A_4: Real Part of $X(1,1)$

A_5: Moment Measure of Skewness

A_6: Pearson's 2nd Measure of Skewness

Here X is the cross-wavelet coefficient matrix, s and t are respectively number of rows and columns in matrix$X$.

VI. SUPPORT VECTOR MACHINE

Although the concept of SVM came up as early as late 70s [11], it was firmly established in 1995 by Cortes and Vapnik. The basic problem that drove the development of SVM was the idea of striking a right balance between the classifier's training performance on a finite amount of training dataset and its generalization ability [12]. SVM is a supervised classifier whose primary aim to solve a binary classification problem by formulating the learning problem as a quadratic optimization problem which has no local minimum and has global optimum [13]. The main advantage of the problem lies in the fact that it scales with the training set size rather than the feature space dimension [14]. SVM classifies two-class data by creating an Optimal Separating Hyperplane (OSH) using the Structural Risk Minimization (SRM) principle [15]. For a linearly separable training dataset there can possibly be infinite many separating hyperplane that will accurately separate the training dataset. The optimal separating hyperplane in this case is the one that correctly classify all the training set vectors and the distance between the hyperplane and the closest vector is maximum[16]. The optimality condition of the separating hyperplane ensures that SVM generalizes well. Fig 4 illustrates how an hyperplane optimally separates two linearly separable classes .X axis represents feature 1 and Y axis represents feature 2

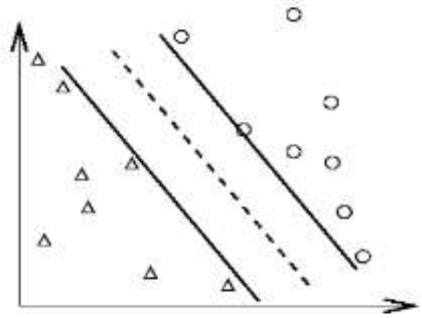

Fig 4: How a hyperplane optimally separates two linearly separable classes

Let us assume a linearly separable training dataset $\{(x_i, y_i)\}[i = 1\; to\; n]$ where x_i are the n-dimensional feature vectors and Y_i are the target/class variables corresponding to them. The Features vectors having corresponding target output y=1 belongs to the positive class and those having corresponding target output y = -1 belongs to the negative class. The decision hyperplane can be represented as:

$$W.x + b = 0 \quad (2)$$

such that:

$$W.x_i + b \geq 1\; for\; y_i = 1 \quad (3)$$

$$\text{and } W.x_i + b \leq -1\; for\; y_i = -1 \quad (4)$$

This can be generalized as:

$$y_i * (W.x_i + b) \geq 1 \;\forall\, i \quad (5)$$

The optimal separating hyperplane is defined as the one that maximizes the distance between projections of the training vectors of the two different classes. The distance ρ can be expressed as a function of hyperplane parameters W and b as:

$$\rho(\mathrm{W}, \mathrm{b}) = \min_{[x:y=1]} \frac{\mathrm{x.W}}{|\mathrm{W}|} - \max_{[x:y=-1]} \frac{\mathrm{x.W}}{|\mathrm{W}|} = \frac{2}{|W|} \quad (6)$$

As the OSH is to maximize ρ for a given set of training data, we can define the hyperplane to be Optimal hyperplane for which $\frac{1}{2}|\mathrm{W}|^2$ is minimized. So, the optimization problem for SVM in case of linearly separable training dataset can be expressed as:

$$\text{Minimize } \phi(\mathrm{W}) = \frac{1}{2}|\mathrm{W}|^2 \quad (7)$$

$$\text{Subject to } y_i * (W.x_i + b) \geq 1 \;\forall\, i \quad (8)$$

However most of the practical problems encountered are not linearly-separable. In such cases, one would like to separate the training case with minimal no. of errors. To express this mathematically, some non-negative variables $\xi_i > 0$ is used. We can now minimize the function:

$$\phi(\xi) = \sum_{i=1}^{n} \xi^{\alpha} \quad (9)$$

Subject to constraints

$$y_i * (W.x_i + b) \geq 1 - \xi_i \;\forall\, i \quad (10)$$

$$\xi_i \geq 0 \;\forall\, i \quad (11)$$

The function $\sum_{i=1}^{n} \xi^{\alpha}$ represents the number of training errors for sufficiently small α. Hence, minimizing it results into minimization of training errors [17]. So, for non-linearly separable cases we obtain the soft-margin hyperplane by minimizing the functional

$$\phi(W, \xi) = \frac{1}{2}|\mathrm{W}|^2 + \mathrm{C}\left(\sum_{i=1}^{n} \xi^{\alpha}\right)^k \;, k > 1 \quad (12)$$

Subject to constraints

$$y_i * (W.x_i + b) \geq 1 - \xi_i \;\forall\, i \quad (13)$$

$$\xi_i \geq 0 \;\forall\, i \quad (14)$$

In case of constructing a polynomial of degree K in an n-dimensional input space, one has to construct an n^k dimensional feature space and then construct the hyperplane in it[18]. This results in very high computational cost. However, It was proven in 1992 [19], that if we take dot product or some other distance measure of the support vector and the input vectors in input space and then use some nonlinear transformation(Kernel functions) to project that into feature space, the hyperplane construction works properly at reasonable computational cost. Some popular Kernel functions are RBF kernel, linear kernel, polynomial kernel etc.

VII. Results

We here use RBF kernel and use Grid Search to obtain optimized value for Kernel Parameters C and γ. The RBF kernel can be expressed as:

$$K(x, x') = e^{-\frac{|X - X'|^2}{-2\sigma^2}} \quad (15)$$

RBF kernel projects the input vectors to an infinite dimensional feature space where they can be classified by linear hyperplane. We use exponentially growing values of C and γ to execute the grid search to find the best parameters for training. It has been found that this is a practical method to find good parameters [20].

We have used 10-fold Cross Validation to obtain a measure of accuracy of the classification. In modern day, although we are capable of computing up to several folds but here the ten-fold stratified cross validation is perhaps the best method for real world datasets [21]. We run the Grid Search for 20 times and obtain the average accuracy to deal with the randomness due to random allocation of dataset in 10-fold cross validation. We obtain the maximum value of accuracy 84.90% at the parameter values C = 1 and γ = 2^2.1. Table 1 shows classification accuracy of our proposed classifier for different C and γ. All the accuracies are for test case.

Classification Accuracy (%)			
C	**γ**		
	2^2	***2^2.1***	***2^2.2***
1	84.82	**84.90**	84.42
2^0.5	84.42	84.55	84.40
2^1	84.10	84.02	83.72
2^1.5	84.35	84.35	83.92
2^2	83.92	84.20	84.15
2^2.5	83.22	83.99	84.50
2^3	82.64	82.86	83.67
2^3.5	81.93	82.16	82.71
2^4	81.38	81.28	81.66
2^4.5	81.51	81.46	81.18
2^5	81.63	81.26	81.16
2^5.5	81.08	81.11	81.18
2^6	80.88	80.75	80.90
2^6.5	79.92	80.55	80.48
2^7	80.05	79.72	80.05
2^7.5	80.35	79.72	79.52
2^8	80.48	80.25	79.47
2^8.5	80.43	80.43	80.10
2^9	80.50	80.25	80.23
2^9.5	80.50	80.43	80.43
2^10.5	80.50	80.43	80.38
2^11	80.50	80.43	80.38
2^11.5	80.50	80.43	80.38
2^12	80.50	80.43	80.38
2^12.5	80.50	80.43	80.38
2^13	80.50	80.43	80.38

Table1. Classification Accuracy of the classifier

VIII. Conclusion and Future Work

Crosswavelet Transform has not been used before for the classification of epilepsy from EEG signals.Through our work we have attempted to develop a general purpose scheme by cross-wavelet transform method in order to find whether the person is suffering from epilepsy or not. The scheme can be further used to classify all the sets and hence develop a general purpose scheme using this method. In order to diagnose epilepsy, the EEG report of the individual along with some additional clinical information has to be gathered thereby proving it to be not an easy task indeed. Hence this type of classifier can be very helpful for taking decisions of whether a person is epileptic or not. Using the steps shown in section 3 we have achieved accuracy as high as 84.90%. To the best of our knowledge none of the literature reported so far, the classification of EEG signals using dataset B and E has attained this level of accuracy. Future work can be done by classifying all the 5 sets of the dataset instead of binary classification.

References

[1]. Subasi, Abdulhamit. "EEG signal classification using wavelet feature extraction and a mixture of expert model." *Expert Systems with Applications* 32.4 (2007): 1084-1093.

[2]. Chandaka, Suryannarayana, Amitava Chatterjee, and Sugata Munshi. "Cross-correlation aided support vector machine classifier for classification of EEG signals." *Expert Systems with Applications* 36.2 (2009): 1329-1336

[3]. Banerjee, Swati, and Madhuchhanda Mitra. "Application of cross wavelet transform for ECG pattern analysis and classification." IEEE transactions on instrumentation and measurement 63.2 (2014): 326-333.

[4]. Subasi, Abdulhamit, et al. "Neural network classification of EEG signals by using AR with MLE preprocessing for epileptic seizure detection."*Mathematical and Computational Applications* 10.1 (2005): 57-70.

[5]. Andrzejak, Ralph G., et al. "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state." *Physical Review E* 64.6 (2001): 061907.

[6]. Chandaka, Suryannarayana, Amitava Chatterjee, and Sugata Munshi. "Cross-correlation aided support vector machine classifier for classification of EEG signals." *Expert Systems with Applications* 36.2 (2009): 1329-1336

[7]. Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.

[8]. http://www.glaciology.net/wavelet-coherence

[9]. D Dey, B Chatterjee, S Chakravorti, S Munshi. "Cross-wavelet transform as a new paradigm for feature extraction from noisy partial discharge pulses." IEEE Transactions on Dielectrics and Electrical Insulation, 17 (1), 157-166

[10]. D Dey, B Chatterjee, S Chakravorti, S Munshi. "Rough-granular approach for impulse fault classification of transformer using cross-wavelet transform." IEEE Transactions on Dielectrics and Electrical Insulation 15 (5), 1297-1304

[11]. V. Vapnik. *Estimation of Dependences Based on Empirical Data [in Russian]*. Nauka, Moscow, 1979.(English translation: Springer Verlag, New York, 1982)

[12]. Burges, Christopher JC. "A tutorial on support vector machines for pattern recognition." *Data mining and knowledge discovery* 2.2 (1998): 121-167.

[13]. Begg, Rezaul K., Marimuthu Palaniswami, and Brendan Owen. "Support vector machines for automated gait classification." *IEEE Transactions on Biomedical Engineering* 52.5 (2005): 828-838.

[14]. Fine, Shai, and Katya Scheinberg. "Efficient SVM training using low-rank kernel representations." *Journal of Machine Learning Research* 2.Dec (2001): 243-264.

[15]. V. Vapnik. *Estimation of Dependences Based on Empirical Data [in Russian]*. Nauka, Moscow, 1979.(English translation: Springer Verlag, New York, 1982)

[16]. apnik, Vladimir N. "An overview of statistical learning theory." *IEEE transactions on neural networks* 10.5 (1999): 988-999.
[17]. Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
[18]. Vapnik, Vladimir N. "An overview of statistical learning theory." *IEEE transactions on neural networks* 10.5 (1999): 988-999.
[19]. Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.
[20]. Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1-16.
[21]. Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *Ijcai*. Vol. 14. No. 2. 1995.