Fuzzy entropy for Feature optimization In Motor Imagery based Brain Computer Interface

Vrushali Raut
E&TC Dept.
Sinhgad College of Engineering
Pune, India
vgraut.scoe@sinhgad.edu

Sanjay Ganorkar
E&TC Dept.
Sinhgad College of Engineering
Pune, India
srganorkar.scoe@sinhgad.edu

***Abstract*— In non-invasive Motor Imagery (MI) based Brain Computer Interface, variation due to MI has spread not only in time domain but also in frequency domain. Even channels are also occupied by this spread. Thus number of features belonging to all these variations is responsible for classifying the underlying task. This paper works on feature optimization using fuzzy entropy so as to avoid under as well over fitting of classifier. Time-Frequency correlation of the signal is obtained using wavelet transform. Second and third order statistical features are extracted from wavelet bands. SVM and KNN with kernel variations are used for classification. Outcome of this experimenting leads to accuracy of 93.7% for optimized features using fuzzy entropy compared to less than 90% for features without optimization.**

Keywords—Motor Imagery (MI), Brain Computer Interface (BCI), Fuzzy Entropy

I. INTRODUCTION

Communication as well control between brain and computer is symbolizes by Brain-Computer Interface(BCI). Electrical activity of the brain collected via electroencephalographic signal can be used non-invasively for such communication. This type of medium will prove to be very useful for disables. Basically BCI can be categorized as evoked and spontaneous. There are two types of BCI: spontaneous EEG, and evoked EEG based BCI systems. Specific mental activity such as motor imagery and related change in brain signal, falls under the category of former Motor Imagery(MI) is one of the example[1]. Whereas neural stimulation from outside as in evoked potential comes under later category, steady state visually evoked potential(SSVEP) is the example of this category[2]. MI terms to be one of the efficient inputs to BCI providing with number of distinguishable movements[3].

The associated signals collected from the electrodes placed on the scalp are weak as well contaminated not only by neighbouring EEG signals but also by other physiological signals like electro-cardiogram, electromyogram etc. This reason proves to be sufficient for use of efficient signal processing at every stage such as to pre-processing or signal enhancement, for feature extraction and to classify the signals[4]. Even the in between steps for feature selection and optimizations are included. It has been noticed that the conventional methods used for signal enhancement are Common Average Referencing[5], Independent Component Analysis, Principal Component Analysis[6], Kalman filtering, Weiner filtering etc. These all methods deal with filtering and boosting the signal. Surface Laplacian calculates the potential belonging to the particular electrode, using the potential on all other electrodes as well shape of the scalp. The derived potential can be considered to be reference independent[7].

Feature extraction methods covered in literature ranges from time domain to frequency domain to mixed domain methods. Motor imagery can be detected from event related synchronization (ERD) and event related de-synchronization (ERD) of the EEG signal. ERD can be captured from the electrode on ipsilateral side of brain, in mu band (8-12Hz), whereas ERS can be captured from the contralateral side in beta band(15-30Hz)[8]. These characteristic of MI signal leads to the preference of the frequency domain methods like Fourier transform(FT)[9], further time frequency correlation and localization can be established using Short time Fourier transform(STFT), Wavelet transform as well Wavelet packet transform which are mixed domain methods also got inclination[10].

Signal for BCI has spatial spread on the electrodes, selecting the features specific to the modulation due expected signal is thus the important task which can be covered under feature reduction or optimization. Principal Component analysis and Independent Component Analysis are the methods preferred for BCI according to literature. These methods help to select principal components from the feature set, or to separate independent components thus reducing redundancy of the features. Referring other biomedical signals and database it has been found that the methods like fuzzy entropy can be effectively used for feature selection as in Parkinsons and dermatology data sets[11].

Classification is one of the challenging parts of signal processing to be dealt with in BCI. Though the variation and related bands are known to extract the features from, but they are completely subject dependent it is not that straight forward to use the trained model of one subject to classify the task of other subjects. Literature suggested supervised learner like Support Vector Machine(SVM)[12] and K Nearest Neighbour(KNN) for classification. Unsupervised Neural networks works well with BCI[13].

II. WAVELET TRANSFORM FOR TIME FREQUENCY ANALYSIS

For MI, EEG signals can be described in terms of Event related synchronization (ERS) and Event related de-synchronization (ERD). These are time as well frequency dependent signals, belonging to 'beta' and 'mu' band respectively and occurring before mechanical movement. The method to be selected for extracting features from these signals must have correlation capacity of these two domains. Thus wavelet transform(WT) seems to be obvious choice, which can extract time as well frequency related information. Implementation of wavelet transform is possible using discrete wavelet transform using "(1)"

$$\psi_{j,k}(t) = a_0^{-j/2}\psi(a_0^{-j}t - kb_0) \quad (1)$$

Where j is scaling parameter and k is translation parameters. A common choice for a0 and b0 are 2 and 1 respectively, which lend them to dyadic sampling grid with wavelet function as in "(2)".

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \quad (2)$$

Wavelet basis selection can further add to the advantages of WT as the matching wavelet can mimic the signal with more accuracy. In this work db10 and bior6.8 are two empirically selected matching wavelets on basis of energy in approximation band. Sampling frequency of the database used is 100Hz, whereas ERD and ERS occurs in the frequency band of 8-12Hz and 15-30Hz thus levels of wavelet decomposition has been computed as suggested by "(3)". Decomposed band accommodate the frequency range of interest in approximate and detail bands as shown in "table I".

$$2^{-j-1}F_s < \Delta F_j < 2^{-j}F_s \quad (3)$$

Where F_s is sampling frequency which is 100Hz in this case and j is level of decomposition. Extracting the wavelet coefficients from the wavelet bands gives precise representation of the signal. It can be directly used as the features or the statistical components can be extracted from it for compress representation of the signals. This work suggested use of second order and third order statistical features for concise representation suitable for classification. Skewness and kurtosis are third order statistical features helpful for representation of dynamics of the signal[14].

TABLE I BAND DETAILS CORRESPONDING TO WAVELET DECOMPOSITION

Decomposed signal	Frequency Range(Hz)
D1	51-100
D2	26-50
D3	12.5-25
A4	0-12.5

III. FUZZY ENTROPY

Estimating level of participation of features for accurate classification of the task will leads to optimisation of features. It reduces the probability of misclassification and thus plays very important role in classification. Transforming the signal in relevant features is to be followed by optimization. Entropy is measures of relevant information in the signal. Fuzzy entropy further indicates the grade of membership of the features for particular class. This can help to decide the importance of the feature for classification. The principle of segmentation of features depending on it's grade or membership value for particular class, fuzzy entropy is implemented in this paper for MI based BCI. It was previously utilized for region growing in aerial images as well in biomedical data such as magnetic resonance spectra[15]. It can be explained as low level segmentation method using concept of fuzzy clustering. The main assumption made here is that the measurement vector belonging to same class cluster together whereas those belonging to different classes lie apart in measurement space. Extracted feature vectors are allowed to belong to various classes with different degrees or membership using fuzzy set. Fuzzy set has capability to model imprecision in non-statistical manner. Quantification of this fuzziness can be helpful in deciding on resolution uncertainty of feature vector. Based on the degree or membership value one can decide importance of the feature for that class. Fuzzy c mean clustering can be used for calculating the fuzzy entropy for the features.

IV. METHODOLOGY

Under this chapter the sequence of operations performed on the signals are given along with the description of database. Wavelet transform of the signals from selected channels using db10 and bior6.8 wavelet has been taken. Three level of decomposition is offered based on sampling frequency as describe in chapter II. Features are extracted from all decomposed wavelet bands. These features are used for classification without optimization giving classification accuracy for identified task. In second experimenting features are optimized based on values of fuzzy entropy calculated using fuzzy c mean clustering.

A. Detailing of Database:

Data set provided by Intelligent Data Analysis Group, Berlin, Department of Neurology, Neurophysics Group. Recording was done on normal subjects without feedback. The task executed by the subject was to press the keys in a self-chosen order with index or small finger of either right hand or left hand. Typing speed was of 1 key per second[16].

B. Format of the data:

Given are 416 epochs of 500 ms length each ending 130 ms before a keypress. 316 epochs are labeled (0 for upcoming left hand movements and 1 for upcoming right hand movements), the remaining 100 epoches are un-labeled for competition purpose. Data are provided in the original 1000 Hz sampling and in a version downsampled at 100 Hz (recommended)[16].

C. Technical details:

The recording was made using a Neuro-Scan amplifier and a Ag/AgCl electrode cap from ECI. 28 EEG channels were measured at positions of the international 10/20-system (F, FC, C, and CP rows and O1, O2). Signals were recorded at 1000 Hz with a band-pass filter between 0.05 and 200 Hz[16].

D. Fuzzy C Mean Clustering(FCM)

Fuzzy entropy method is based on the utilization of the Fuzzy C-Means Clustering algorithm[17]. FCM is used to construct the membership function of all features. The data may belong to two or more clusters simultaneously and the belonging of a data point to the clusters is governed by the membership values. Similar data points are placed in the same cluster and dissimilar data points normally belong to different clusters. The membership values of the data points are reorganized iteratively to reduce the dissimilarity. The Euclidean distance is used to measure the dissimilarity of two data points.

The FCM algorithm is explained as follows

Step 1: Assume the number of clusters (C), where 2<C<N, C – number of clusters and N – number of data points.

Step 2: Calculate the jth cluster center Cj using the following expression

$$Cj = \frac{\sum_{i=1}^{N} \mu_{ij}^{g} x_{ij}}{\sum_{i=1}^{N} \mu_{ij}^{g}} \pi r^2 \quad (4)$$

Where g>1is the fuzziness coefficient and μij is the degree of membership for the ith data point xi in cluster j.

Step 3: calculate the Euclidean distance between the i^{th} data point and the j^{th} cluster center as follows

$$d_{ij} = |C_j - x_i| \quad (5)$$

Step 4: Update the fuzzy membership values according to dij if dij >0, then

$$\mu = \frac{1}{\sum_{m=1}^{C} \left(\frac{d_{ij}}{d_{im}}\right)^{\frac{2}{g-1}}} \quad (6)$$

If d=0, then the data point coincides with the jth cluster center (C) and it will have the full membership value, i.e., μij =1.0

Step 5: repeat Steps 2–4 until the changes in μ are less than some pre-specified values.

The FCM algorithm computes the membership of each sample in all clusters and then normalizes it as in equation 7.

$$CD_c(A) = \frac{\sum_{x\epsilon c} \mu_A(x)}{\sum_{x\epsilon c} \mu_A(x)} \quad (7)$$

The fuzzy entropy FEc(A) of class c is defined in equation 8

$$FE_c(A) = -CD_c(A) \log_2 CD_c(A) \quad (8)$$

The fuzzy entropy FE (A) of a fuzzy set X is as given by equation 9

$$FE(A) = \sum_{x \epsilon C} FE_c(A) \quad (9)$$

Signal to be classified consist of two variations belonging to left hand movement and right hand movement thus one can consider two classes for organizing entropy of the signal. Fuzzy entropy is calculated for the training feature belonging to left hand movement and right hand movement. Features having less value of entropy are selected. In this manner optimized/reduced feature set is constructed separately for two movements. This feature set is used to train the SVM and KNN classifier and tested for various kernel functions.

V. Results

Wavelet basis functions used are db10 and bior6.8, which are applied on the signals from channel c4 to extract wavelet coefficients. Statistical coefficients are extracted from all the wavelet bands and are passed to SVM and KNN classifier.

Table II gives classification accuracy for the features without optimization which is 84% maximum. Whereas Table III depicted classification accuracy of 93.7% with optimized features using fuzzy entropy.

Table II Percent Classification Accuracy without Optimization

Classifier	Wavelet Basis	
	Db10	**Bior6.8**
SVM fine Gaussian kernel	**85**	**81**
SVM Cubic kernel	**83.5**	**80**
KNN weighted	**84**	**83.5**

Table III Percent Classification Accuracy with Optimization

Classifier	Wavelet Basis	
	Db10	**Bior6.8**
SVM fine Gaussian kernel	**92**	**87.4**
SVM Cubic kernel	**91.3**	**88**
KNN weighted	**93.7**	**90.6**

VI. Conclusion

This works promote wavelet transform for time-frequency representation of the signal. As the expected modulation due to MI is frequency band limited, three level of wavelet decomposition is proposed in this paper. Higher order statistical features used in this work represent dynamics of the signal which can further leads to good classification accuracy. Fuzzy entropy proposed for feature reduction/optimization increases classification accuracy from 84% to 93.7%.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," Clin. Neurophysiol., vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "SSVEP recognition using common feature analysis in brain-computer interface," J. Neurosci. Methods, vol. 244, pp. 8–15, 2015.

[3] G. Prasad, P. Herman, D. Coyle, S. McDonough, and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: A feasibility study," J. Neuroeng. Rehabil., vol. 7, no. 1, pp. 1–17, 2010.

[4] S. G. Mason and G. E. Birch, "A general framework for brain - Computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 1, pp. 70–85, 2003.

[5] P. Speller and M. J. Alhaddad, "Common Average Reference (CAR) Improves Common Average Reference (CAR) Improves P300 Speller," *Int. J. Eng. Technol. Vol.*, vol. 2, no. March, 2012.

[6] T. Wang and B. He, "An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain-computer interface," *J. Neural Eng.*, vol. 1, no. 1, pp. 1–7, 2004.

[7] C. E. Tenke and J. Kayser, "Surface Laplacians (SL) and phase properties of EEG rhythms : Simulated generators in a volume-conduction model," Int. J. Psychophysiol., vol. 97, no. 3, pp. 285–298, 2015.

[8] B. Xu and A. Song, "Pattern recognition of motor imagery EEG using wavelet transform," J. Biomed. Sci. Eng., vol. 1, no. May, p. 64, 2008.

[9] M. E. Elhalawani and A. S. Tolba, "Brain – Computer Interaction (BCI) Using a Fast Algorithm for Empirical Mode Decomposition (EMD) and a Specialized Hardware," Int. J. Inf. Sci. Intell. Syst., vol. 3, no. 1, pp. 169–178, 2014.

[10] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 1, pp. 83–88, 1999.

[11] P. Luukka, "Expert Systems with Applications Feature selection using fuzzy entropy measures with similarity classifier," Expert Syst. Appl., vol. 38, no. 4, pp. 4600–4607, 2011.

[12] P. Herman, G. Prasad, S. Member, T. M. Mcginnity, and D. Coyle, "Comparative Analysis of Spectral Approaches to Feature Extraction for EEG-Based Motor Imagery Classification," vol. 16, no. 4, pp. 317–327, 2008.

[13] R. Chai et al., "Brain – Computer Interface Classifier for Wheelchair Commands Using Neural Network With Fuzzy Particle Swarm Optimization," IEEE J. Biomed. Heal. INFORMATICS, vol. 18, no. 5, pp. 1614–1624, 2014.

[14] Y. Kutlu and D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," Comput. Methods Programs Biomed., vol. 105, no. 3, pp. 257–267, 2011.

[15] D. Li, W. Pedrycz, and N. J. Pizzi, "Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification," IEEE Trans. Biomed. Eng., vol. 52, no. 6, pp. 1132–1139, 2005.

[16] B. Blankertz et al., "Classifying Single Trial EEG : Towards Brain Computer Interfacing," Adv. Neural Inf. Proc. Syst. 14 (NIPS 01), no. c, pp. 96–98, 2002.

[17] K. V. Bulusu and M. W. Plesniak, "Shannon entropy-based wavelet transform method for autonomous coherent structure identification in fluid flow field data," Entropy, vol. 17, no. 10, pp. 6617–6642, 2015.