Research Journal of Applied Sciences, Engineering and Technology 10(9): 1035-1037, 2015
DOI: 10.19026/rjaset.10.1871
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: March 12, 2015 Accepted: April 1, 2015 Published: July 25, 2015

Research Article

Fuzzy Mutual Information as a Dimensionality Reduction Technique for Epileptic Electroencephalography Signals

R. Harikumar and P. Sunil Kumar
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India

Abstract: The aim of this study is to use Fuzzy Mutual Information as a Dimensionality Reduction Technique for Epileptic Electroencephalography Signals. To design an effective classification model, it is vital to extract a small set of closely related relevant features from a data set which has a high dimension. Such a type of procedure should explore the series of estimations of the relationship between each and every pair of variables. Also, the estimation is done between the two variables and also for the class labels too. For continuous and hybrid data there are various other strategies that are useful for the estimation of mutual information. Fuzzy Mutual Information is very helpful for obtaining the most stable feature sets and the relationships between two variables is accurately estimated. In this study, Fuzzy Mutual Information is applied as the dimensionality reduction technique for the electroencephalography signals obtained from epileptic patients. The Electroencephalogram (EEG) is actually a measure of the cumulative firing of neurons in various parts of the brain. The EEG contains the information with regard to the changes in the electrical potential of the brain which is obtained from a set of recording electrodes. Here the results are discussed using Fuzzy Mutual Information technique as a dimensionality reduction technique for the processing of electroencephalography signals from an epileptic patient.

Keywords: EEG, epilepsy, fuzzy mutual information

INTRODUCTION

The Electroencephalogram (EEG) is actually a measure of the cumulative firing of neurons in various parts of the brain (Kumar and Agnihotri, 2010). The obtained data includes both the standard waveforms and shortly occurring electrical patterns. The standard waveforms may accompany with rapid variations in amplitude, frequency and phase. The electrical patterns such as sharp and spike waveforms and spindles may also be present. EEG patterns can be modified using a wide range of variables including hormonal, biochemical, metabolic, circulatory, neuro electric and behavioral factors (Iasemidis *et al*., 2003). Earlier, just by visual inspection the encephalographer was able to distinguish the normal EEG activity from the abnormal EEG activity. The most important activity detected from the EEG is the epilepsy and it is characterized by the excessive activity by a part or all of the central nervous system. By observing the different EEG waveform patterns the different types of epileptic seizures are characterized (Van Gils *et al*., 1997). In order to quantify the changes occurring based on the EEG signals, the application of computers has made it possible to effectively apply a host of methods for the real time monitoring and detection of epileptic seizures. The EEG is a vital tool used for the diagnosis, monitoring and managing of neurological disorders related to epilepsy (Patnaik and Manyam, 2008).

With a drastic improvement in detection and storage technologies which are developing, for the purpose of pattern recognition and training machine learning modules a much greater volume of data has been produced. However, most of the features in the data are usually either irrelevant to the obtained classes or redundant. The high dimensionality also poses a great challenge to problems in the field of pattern recognition, data mining and machine learning (Acharya *et al*., 2013). Therefore, selection if a small set of importance and useful features with the consideration of a minimum redundancy in between the features and a class label is an important process for the further design of an effective classification system. Mutual information has a lot of advantages and it is widely used as both a feature extraction and dimensionality reduction technique (Sanchez *et al*., 2005). The greatest use of using mutual information is that it is used for the estimation of the dependency level between two variables. Also Mutual Information is used to measure more complicated non-linear relationships along with the general measurement of the linear (Pearson's correlation) and monotonic relationships (Spearman's correlation). Initially mutual information was used for the estimation purposes to check the dependency between two discrete variables. Also

Corresponding Author: R. Harikumar, Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India

Mutual Information was applied to the estimation of the continuous variables with suitable additional transformations. The input data has a lot of noisy characteristics and therefore the incorporation of the fuzzification processes should happen which would make the process more noise-tolerant. Therefore Mutual Information has a lot of uses and it can be used widely for various types of applications in biomedical field. In this study the Fuzzy Mutual Information is used as a Dimensionality Reduction technique for the epileptic electroencephalography signals.

MATERIALS AND METHODS

The raw EEG data of an epileptic patient in European Data Format (EDF) who was under treatment in the Neurology Department of Sri Ramakrishna Hospital, Coimbatore was obtained. An issue that has been given great attention is the preprocessing stage of the EEG signals because it is important to use the best technique to extract the useful information embedded in the non-stationary biomedical signals. The EEG records which was obtained were continuous for about 30 sec, each of them were divided into epochs of approximately two second duration. To detect any significant changes in activity and presence of artifacts and also short enough to avoid any redundancy in the signal a two second epoch is long enough. For a patient we have 16 channels over three epochs. In this study we have reported the dimensionally reduced values using Fuzzy Mutual Information Technique for all the sixteen channels under three epochs. Having a frequency of 50 Hz, each epoch was sampled at a frequency of 200 Hz. Each sample corresponds to the instantaneous amplitude values of the signal, totaling to 400 values for an epoch. We choose a sample window of 400 points corresponding to 2 sec of the EEG data. This particular width can cover almost all types of transient epileptic patterns in the EEG signal, even though Seizure can often last longer.

ESTIMATION OF FUZZY MUTUAL INFORMATION

Estimation of mutual information: Mutual Information can be estimated in both the discrete data form and the continuous data form.

Estimation in discrete data: For the discrete data, the mostly used Mutual Information method is based on the Shannon entropy. For a given system Q with Z_Q possible states, the Shannon entropy is given as follows:

$$S(Q) = -\sum_{i=1}^{Z_Q} r(q_i) \log r(q_i)$$

where,

$$Q = q_i \, (1 \leq I \leq Z_Q)$$

Similarly the joint entropy can be computed as follows:

$$S(Q,W) = -\sum_{j=1}^{Z_Q} \sum_{j=1}^{Z_W} r(q_{i,} w_j) \log r(q_{i,} w_j)$$

The Mutual Information I (Q, W) between these two systems Q and W is defined as follows:

$$I(Q,W) = S(Q) + S(W) - S(Q,W)$$

Estimation in continuous data: For the estimation of the mutual information involved with a series of probability distribution between two continuous variables, always a discretization step of the data with an uniform binning is usually performed. The global binning methods are most widely used for the partitioning of a continuous variable A into B discrete bins. The computation of the mutual information between the given continuous variables based on the probability distribution could be easily carried out. This kind of method is very simple and efficient, because in the real data, the distribution of each continuous variable is varied and another additional normalization process would be required. The precision of the resulting MI is determined by the appropriate setting of discrete bins which are usually targeted.

Estimation of fuzzy MI: Instead of discretizing the data which is a continuous form, fuzzifying the data before the computation of mutual information seems to be a versatile measure. Between a categorical variable and a fuzzy variable the natural definition of the mutual information is obtained as a fuzzy number. Therefore only Fuzzy Mutual Information came into existence (Sanchez *et al*., 2005). In FMI for the measurement purposes, B-Spline function are generally used. The B-spline method is more time efficient and it is analogous to the kernel density estimators. It is generally difficult from the classical binning method. It allows each data point to be assigned to several bins simultaneously at the same time.
The Fuzzy MI is defines as follows:

$$FH(R_L) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{\left|[x_i]R_L\right|}{n}$$

where,

$$\left|[x_i]R_L\right| = \sum_{i=1}^{n} r_{ij}$$

If two subsets L, L_1 and L_3 are given, the fuzzy joint information entropy is given as follows:

Table 1: FMI, hurst exponent for FMI and CTM for FMI values for channel 1

	Channel 1		
Parameter	Epoch 1	Epoch 2	Epoch 3
Fuzzy mutual information values	0.3456	0.7151	0.5648
Hurst exponent for FMI values	0.4200	0.3200	0.4300
Centre tendency mode for FMI values	0.4118	0	0

Table 2: Average FMI, hurst exponent for FMI and CTM for FMI values for an epileptic patient

Parameter	Avg. value for 16 channels
Fuzzy mutual information values	0.6317
Hurst exponent for FMI values	0.3900
Centre tendency mode for FMI values	0.1372

Avg.: Average

$$FH(L_1.L_2) = FH(R_{L1,}R_{L2}) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{\left|[x_i]_{F_1} \cap [x_i]_{F_2}\right|}{[x_i]_{F_1}}$$

The Fuzzy MI between L_1 and L_2 is denoted as follows:

$$FMI(F_1;F_2) = FH(F_1) + FH(F_2) - FH(F_{1,}F_2)$$

Based on the crisp equivalence relation, it has been demonstrated or understood that Shannon's entropy is more identical to fuzzy information entropy. For the estimation of mutual information between any two continuous variables, discretization or fuzzification of the data is not necessary. Therefore it is the best method for us to analyze the continuous data by the estimation analysis of fuzzy mutual information. To measure the similarity between two continuous variables x and y, the following relation is used and expressed as follows:

$$C(x,y) = \exp(-\|x-y\|)$$

RESULTS AND DISCUSSION

The following results show the Dimensionality reduced values of the EEG Signals of Epileptic Patient. For the first channel and under three epochs the values of Fuzzy Mutual Information is computed. Table 1 shows the Fuzzy Mutual Information, Hurst Exponent for Fuzzy Mutual Information and Centre Tendency Mode for Fuzzy Mutual Information Values for Channel 1. Table 2 shows the average values of Fuzzy Mutual Information, Hurst Exponent for Fuzzy Mutual Information value and the Centre Tendency Mode for Fuzzy Mutual Information value.

The Hurst Exponent Values are taken for the Fuzzy Mutual Information Values and it is found to be approximately below the range of 0.5 which clearly exhibits the non-linearity nature and the rhythmicity of the EEG Signals. Also when the Centre Tendency mode for FMI values were computed for channel 1 of the patient, only under epoch 1 it produced a value of about 0.4118 whereas for epoch 2 and 3 it produced a value of zero only. Future work is planned for the classification of epilepsy risk level from EEG signals using different types of classifiers.

REFERENCES

Acharya, U.R., S.V. Sree, G. Swapna, R.J. Martis and J.S. Suri, 2013. Automated EEG analysis of epilepsy: A review. Knowl-Based Syst., 45: 147-165.

Iasemidis, L.D., D.S. Shiau, W. Chaovalitwongse, J.C. Sackellares, P.M. Pardalos *et al*., 2003. Adaptive epileptic seizure prediction system. IEEE T. Bio-Med. Eng., 50(5): 616-627.

Kumar, P. and D. Agnihotri, 2010. Biosignal denoising via wavelet thresholds. IETE J. Res., 56(3): 132-138.

Patnaik, L.M. and O.K. Manyam, 2008. Epileptic EEG detection using neural networks and post classification. Comput. Meth. Prog. Bio., 91: 100-109.

Sanchez, L., M.R. Suarez and I. Couso, 2005. A fuzzy definition of Mutual Information with application to the design of Genetic Fuzzy Classifiers. Proceeding of the International Conference on Machine Intelligence. Tozeur, Tunisia.

Van Gils, M., A. Rosenfalck, S. White, P. Piror, J. Gade *et al*., 1997. Signal processing in prolonged EEG recordings during intensive care. IEEE Eng. Med. Biol., 16(6): 56-63.