ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 8, Issue 4, 2015

ISSN - 0974-2441

Research Article

PERFORMANCE ANALYSIS USING CODE CONVERTER APPROACH AND THE APPLICATION OF APPROXIMATE ENTROPY AS POST CLASSIFIER FOR THE CLASSIFICATION OF EPILEPSY RISK LEVELS FROM EEG SIGNALS

HARIKUMAR R[1], SUNIL KUMAR P[2*]

Department of ECE, Bannari Amman Institute of Technology, India. Email: sunilprabhakar22@gmail.com

Received: 08 May 2015, Revised and Accepted: 20 May 15

ABSTRACT

Objective: 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. The aim of this paper is to give a performance analysis by considering the advantage of Code Converter technique and Approximated Entropy (ApEn) is used as a post classifier for the classification of the epilepsy risk levels obtained from EEG signals.

Methods: The Data Acquisition of EEG signals are done initially from the hospital. Then the code converter approach is presented, as working on definite alphabets is much easier when compared to that of working on numericals. Finally, ApEn is used as a Post Classifier for the classification of epilepsy risk levels from EEG signals.

Results: The Performance Index and Quality Values are the two important parameters that are used to assess the performance of the Code Converters and the Classifier. The Perfect Classification rate of 83.94% is achieved along with an Accuracy of 91.97% and a Quality Value of 18.5.

Conclusion: The computation of this procedure seems to be very simple and versatile. Future works may use different Dimensionality Reduction techniques to analyze its performance with Approximated Entropy as Post Classifier.

Keywords: Electroencephalogram signals, Code converter, Performance index, Quality values.

INTRODUCTION

The electroencephalogram (EEG) is a vital tool used for the diagnosis, monitoring and managing of neurological disorders-related to epilepsy [1]. The obtained data include both the standard waveforms and shortly occurring electrical patterns. The standard waveforms may accompany with rapid variations in amplitude, frequency and phase [2]. 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, neuroelectric and behavioral factors [3]. 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 epilepsy, and it is characterized by the excessive activity by a part or all of the central nervous system [4]. By observing the different EEG waveform patterns, the different types of epileptic seizures are characterized. In order to quantify the changes occurring based on the EEG signals, the application of computers has made it possible to apply effectively a host of methods for the real-time monitoring and detection of epileptic seizures. By the sudden recurrent and transient disturbances of mental function which results in the over discharge group of brain cells, it leads to the characterization of epilepsy [5]. The organization of the paper is as follows: Section 1 introduces the paper and materials and methods are discussed in Section 2. Section 3 describes about the analysis of approximated entropy (ApEn) and the results are discussed and concluded in Section 4.

METHODS

Data acquisition of EEG signals

For the performance analysis of the epilepsy risk levels using code converter approach and entropy as a post classifier, the raw EEG data of 20 epileptic patients who were under treatment in the Neurology Department of Sri Ramakrishna Hospital, Coimbatore in European data format are taken for study. The preprocessing stage of the EEG signals is given more attention because it is vital to use the best available technique in literature to extract all the useful information embedded in the non-stationary biomedical signals. The EEG records which were obtained were continuous for about 30 seconds, and each of them was divided into epochs of 2 seconds duration. In general, a 2-second epoch is long enough to avoid unnecessary redundancy in the signal, and it is long enough to detect any significant changes in activity and to detect the presence of artifacts in the signal. For each and every patient, the total number of channels is 16 and it is over three epochs. The frequency is considered to be 50 Hz and the sampling frequency is considered to be about 200 Hz. Each and every sample corresponds to the instantaneous amplitude values of the signal which totals to 400 values for an epoch. The total number of artifacts present in the data is four. Chewing artifact, motion artifact, eye blink and electromyography are the four number of artifacts present and approximately the percentage of data which are artifacts is 1%. No attempts were made to select certain number of artifacts which are of more specific nature. The main objective to include artifacts is to differentiate the spike categories of waveforms from non-spike categories.

The latter could also be a normal background EEG or artifacts. For training and testing the classifiers, a suitable segment of EEG data has to be selected. In our experiment, all the EEG signals were examined visually by a qualified EEG technologist through a short sampling window. As a gold standard, the neurologist's decision regarding EEG features or normal EEG segment was used. A sample window of 400 points which corresponds to 2 seconds of the EEG data was chosen. This particular width was chosen because it can cover almost all types of transient epileptic patterns in the EEG signal. Table 1 shows the parameter representation for various risk levels.

The code converters output is always encoded into the string of seven codes which corresponds to each and every signal parameter, which is entirely based on the epilepsy risk levels threshold values. The noise is present in the form of overlapping ranges in the expert defined threshold values.

Code converter technique

The sampled output values are processed as an individual code with the encoding method [6]. Working on definite alphabets is much easier, whereas processing numbers with a perfect decimal accuracy is more difficult. The representation of the risk level classifications is given in Table 2.

The characteristic representation has a great flexibility and it eases the operation whereas performing the cumbersome operation of numbers is very tedious. By the easy encoding of each risk level in one of the five states, for each of the sixteen channels a string of seven characters is obtained. A sample output with the original patient readings is shown in Fig. 1 for 8 different channels over three epochs.

It is very obvious that a low-risk level can be seen from channel 1 and a high-risk level can be obtained from channel 7. The classification of the risk level also varies between the adjacent epochs. Here, there are totally sixteen different channels as an input to the system at three different epochs. Therefore, the total number of input-output pairs are 48. It is absolutely mandatory to find the exact level of epilepsy risk in the patient since we deal with known cases of epileptic patients. This technique will also help in the development of automated systems that can accurately classify the epileptic patients risk level under observation. In such cases the optimization is absolutely necessary. The EEGer will have a clear picture by improving the classification of the patient. The general outputs from each epoch are not similar and it varies in a condition such as [YYZXXXX] to [WYZYYYY] to [YYZZYYY]. In such cases, the energy factor is highly predominant which results in two epochs having a high-risk level and middle epoch having a low-risk level. The channel five and six settles at a very high-risk level. Because of this type of mixed state output, the grouping of four adjacent channels to optimize the risk level is performed. The patterns, which are frequently repeated, show the average risk level of the channel groups. The constant risk level associated in a particular epoch is being depicted by the same individual patterns. The occurrence of at least one Z pattern in an epoch shows that whether a group of channels is at the high-risk level or not. The variation of the risk level is very abrupt across epochs and it is eventually done in the channels. Hence, a dilemma occurs and the final verdict cannot be easily chosen. Using the weighted positional representation which is shown in Table 3 the five risk levels are encoded as Z>Y>X>W>U in binary strings which have a length of five bits. If each output risk level is encoded, it gives us a string of seven alphabets and the fitness of which is easily calculated as the sum of probabilities of the individual alphabets. For instance, if the output of an epoch is encoded as ZZYXWZZ, then its corresponding fitness would be around 0.419352.

Table 1: Parameter representation for various risk levels

Risk level	Representation
Normal	U
Low	W
Medium	X
High	Y
Very high	Z

Table 2: Representation of risk level classifications

Risk levels					
Normalized parameters	**Normal**	**Low**	**Medium**	**High**	**Very High**
Energy	0-1	0.7-3.6	2.9-8.2	7.6-11	9.2-30
Variance	0-0.3	0.15-0.45	0.4-2.2	1.6-4.3	3.8-10
Peaks	0-2	1-4	3-8	6-16	12-20
Events	0-2	1-5	4-10	7-16	15-28
Sharp waves	0-2	1-5	4-8	7-11	10-12
Average duration	0-0.3	0.15-0.45	0.4-2.4	1.8-4.6	3.6-10
Covariance	0-0.05	0.025-0.1	0.09-0.4	0.28-0.64	0.54-1

Table 3: Binary representation of risk levels

Risk level	Code	Binary string	Weight	Probability
Very high	Z	10000	16/31=0.51612	0.086021
High	Y	01000	8/31=0.25806	0.043011
Medium	X	00100	4/31=0.12903	0.021505
Low	W	00010	2/31=0.06451	0.010752
Normal	U	00001	1/31=0.03225	0.005376
		11111=31	Σ=1	

ApEn as a post classifier

In statistics, ApEn is actually a technique which is used for the quantification of the amount of regularity. Over a time-series data, it is used to quantify the unpredictability of fluctuations also. Based on the various entropy measures, regularity can be originally measured by using exact regularity statistics. To calculate the accurate entropy, it would definitely require vast amounts of data and the results also will be greatly affected by the system noise, therefore for experimental data, it is not practical to apply these methods. Therefore to handle such limitations, a slight modification of an exact regularity statistic parameter was done, and it was termed as ApEn. ApEn has wide range of applications such as the analysis of medical data, remote sensing, psychology and finance.

Algorithm:

Step 1: A time series of data such as a (1), a (2),…., a (N) is done. They are N raw data values obtained from measurement and they are equally spaced in time.

Step 2: A real positive number 'r' and an integer 'm' is fixed. The m represents the run length of the data and the filtering level is specified by 'r.'

Step 3: A sequence of vectors such as a (1), a (2)…. a(N-m+1) in R^m is formed.

Step 4: The sequences a(1), a(2),…. A (N-m+1) is used to construct 'i' where 1<i<N-m+1

Step 5: The following function is then defined as

$$\varphi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \log(C_i^m(r))$$

where log is the natural logarithm for m and r.

Step 6: The appropriate entropy (ApEn) is defined as follows

Epoch1	Epoch2	Epoch3
WYYWYYY	WYYWYYY	WZYYWWW
YZZYXXX	YYYYXXX	YYYXYYY
YYZXYYY	YYYYYYY	YYYYYYY
YZZYXYY	XZZXYYY	YYYYYYY
ZZZYYYY	WYYYXXX	YYYXYYY
YYZXXXX	WYZYYYY	YZZYYYY
ZZZYYYY	YYYYYYY	ZZZYYYY
YYYYXXX	YYYYXXX	YYYXZYY

Fig. 1: Output of code converter for patient 2

$$ApEn = \varphi^{m}(r) - \varphi^{m+1}(r)$$

The ApEn reflects the likelihood that a set of similar patterns and observations will definitely not be followed by additional similar observations. A high predictable process has a less ApEn. The important advantages of ApEn are that it has a very low computational demand. It is always designed to work for very small data samples and it is very useful for the real-time applications. It has a very less effect from noise and its consequences. If the data are very noisy, then the ApEn measure can be easily compared to the actual amount of noise level present in the data in order to determine the quality of true information present in the data with/ without the interference.

RESULTS AND DISCUSSION

For various types of classifiers based on the performance index (PI), quality values (QV), sensitivity, specificity, time and accuracy the results are computed and tabulated in Table 4. The sensitivity and specificity analysis is given in the Fig. 2. The formulae for the PI, sensitivity, specificity, and accuracy are given as follows:

$$PI = \frac{PC - MC - FA}{PC} \times 100$$

where PC - Perfect classification, MC - Missed classification, FA - False alarm,

The sensitivity, specificity and accuracy measures are stated by the following

$$Sensitivity = \frac{PC}{PC + FA} \times 100$$

$$Specificity = \frac{PC}{PC + MC} \times 100$$

$$Accuracy = \frac{Sensitivity + Specificity}{2} \times 100$$

It is inferred from Fig. 2 that the sensitivity is varied throughout and it has a certain amount of abrupt variations and it does not remain constant.

From Fig. 3, it is inferred that the QV shows abrupt variations throughout the time series, and it does not remain constant. When the time delay is approximately 2 seconds, then the obtained QV is high.

From Fig. 4 it is inferred that the accuracy also shows abrupt variations throughout the PI series and it does not remain constant.

The following table summarizes the average values obtained for the 20 patients when code converters act as pre classifier and ApEn act as post classifier.

Thus, the paper gives a performance analysis by considering the code converter technique when ApEn acts as post classifier for the classification of the epilepsy risk levels obtained from EEG signals. PI and QV were the two parameters that were used to assess the performance of the code converter and classifiers. It is concluded that the average perfect classification is 83.94% and the average PI is 78.68%. The average sensitivity and specificity values are 91.66% and 92.28%, respectively. The average time delay computed in seconds is obtained to be 2.142. The average QV obtained in this case is 18.5 and the average accuracy obtained is 91.97%. Future work may incorporate the usage of a variety of dimensionality reduction techniques followed by the classification of epilepsy risk levels using entropy as a post classifier.

Table 4: Average values for all the 20 patients when code converters act as pre classifier and entropy act as post classifier

Parameters	Average values for all the 20 patients (%)
Average perfect classification	83.94
Average PI	78.68
Average sensitivity	91.66
Average specificity	92.28
Average time delay (sec)	2.142
Average QV	18.5
Average accuracy	91.97

QV: Quality value, PI: Performance index

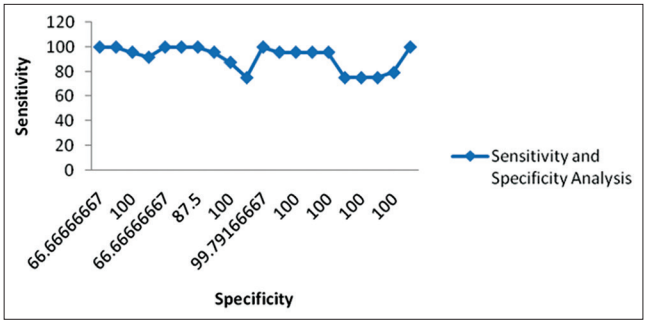

Fig. 2: Sensitivity and specificity analysis using code converter technique when approximated entropy acts as post classifier

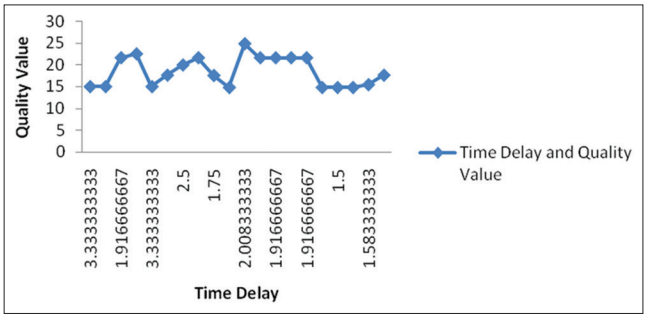

Fig. 3: Time delay and quality value analysis using code converter technique when approximated entropy acts as post classifier

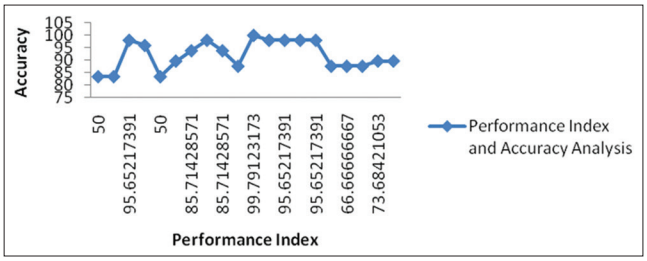

Fig. 4: Performance index and accuracy analysis using code converter technique when approximated entropy acts as post classifier

REFERENCES

1. Niederhauser JJ, Esteller R, Echauz J, Vachtsevanos G, Litt B. Detection of seizure precursors from depth-EEG using a sign periodogram transform. IEEE Trans Biomed Eng 2003;50(4):449-58.
2. Zumsteg D, Wieser HG. Presurgical evaluation: Current role of invasive EEG. Epilepsia 2000;41 Suppl 3:S55-60.
3. Dingle AA, Jones RD, Carroll GJ, Fright RR. A multistage system to detect epileptic form activity in the EEG. IEEE Trans Biomed Eng 1993;40(12):1260-8.

4. Harikumar R, Narayanan BS. Fuzzy Techniques for Classification of Epilepsy risk level from EEG Signals", Proceedings of IEEE Tencon – 2003, Conference on Convergent Technologies for the Asia-Pacific Region Bangalore, India 2003;209-13:14-7.
5. Qu H, Gotman J. A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device. IEEE Trans Biomed Eng 1997;44:115-22.
6. Harikumar R, Sukanesh R, Bharathi PA. Genetic algorithm optimization of fuzzy outputs for classification of epilepsy risk levels from EEG signals. India J Interdiscip Panels 2005;86(1):9-17.