RESEARCH ARTICLE OPEN ACCESS

Epileptic Spike Recognition in Electroencephalogram Using Genetic Algorithm

Shashi Prabha [1], Dr. Tripti Arjarya [2]
Department of Computer Science and Engineering
Bhabha Engineering Research Institute, and Bhopal
MP - India

ABSTRACT

This Paper presents an automated method of Epileptic Spike detection in Electroencephalogram (EEG) using Genetic Algorithm (GA). It takes pre-recorded single channel EEG data file as input and finds the occurrences of Epileptic Spikes data in it. The EEG signal was recorded at 256 Hz in two minutes separate data files using the Visual Lab-M software (AD Link Technology Inc., Taiwan). It was preprocessed for removal of baseline shift and band pass filtered using an infinite impulse response (IIR) Butterworth filter. A system, whose functionality was modeled with GA, has already been proposed. The system was tested with 10 EEG signal data files. The recognition rate of Epileptic Spike as on average was 95.68%. This system does not require any human intrusion. Also it does not need any short of training. The result shows that the application of GA can be useful in detection of different characteristics present in EEG signals. This approach could be extended to a continuous data processing system.

Keywords:- Automatedsystem;Electroencephalogram;Epiltic spike; Genetic Algorithm

I. INTRODUCTION

Complex non linear brain cortical potential or electroencephalogram (EEG) contains variety of information about the state of the patient's heath, thus, establish as a very important none invasive diagnostic tool in the clinical practices. One of the most common pathological alterations in recorded EEG signal can be and the analysis of the complex brain system and its related signals are very tedious jobs and it also takes a great deal of timing as well as experience. However the inter observer variability in analog signal analysis for any nonlinear and non periodic signal like EEG waves are obvious [1-3].

Digital computer, digital signal processing tools and processing systems. Review of literature suggests that a number of automated methods for spike detection exist

to speed up the process. The time domain method was proposed by Liu et al [4].This method searches for periodic rhythmic patterns in EEG similar to ones occurrence during the seizure activity. Furthermore, Gotman et al. [5] suggested a method for the detection of the spike in a given frequency domain by finding the difference in the frequency characteristics of the usual and epileptic EEG.A number of methods to find out the spikes applying obtained from the patient; those are suspected of having epilepsy .the epileptic activity may be started with shape and spontaneous spike pattern and generally known to occur in all age group in all human being since the occurrence of spike pattern can bee considered as the symptom of state of epileptic seizures, hence the direction of these spike plays an important diagnostic role in early predictions of epileptic conditions. The long term analog observation wavelet transform have been given in the past [6-11]. ANN is being widely deployed for the purpose of spike recognition [12-17]. However the use of threshold criteria for the optimization of the automated system was presented in many of the previous works. In addition to this the works using ANN needs the training process initially to train the networks.

1.1 GENETIC ALGORITHM

The genetic algorithm [26] is a search technique used in Computing to find exact or approximate solutions to optimization and search problems. They are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. These are commonly implemented as computer simulations where a population of abstract representations (called chromosomes) of candidate solutions to an optimization problem, evolve towards better solutions. Traditionally,solutions are represented in binary as strings of 0 s and 1 s, but other encodings are also possible. The evolution usually starts from a population of randomlygenerated individuals and gives rise to new generations. In each generation, the fitness of every individual in the population is evaluated based on the solution provided by it. Subsequently, multiple individuals are stochastically selected from the current population (based on their fitness), andmodified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced, a satisfactory fitness level has been reached for the population or when no improvement is seen in the solution to the problem after a certain number of generations. We use a binary representation of the chromosome to indicate whether or not a particular EEG recording is used in the training set. The fitness of the chromosome is then updated based on the value of the seizure detection sensitivity. The algorithm terminates if there is no improvement in the sensitivity value after 200 successive generations

II. THE SPIKE DETECTION PROBLEM

Electroencelograph (EEG) signal plays important role in the diagnosis of epilepsy. The long term EEG recordings of an epileptic patient obtained from the ambulatory recording systems contain a large volume of EEG data. Detection of the epileptic activity requires a time consuming analysis of the entire length of the EEG data by an expert. The traditional methods of analysis being tedious, many diagnostic systems for epilepsy have emerged in recent years. This paper discusses a new diagnostic method (genetic algorithm) for epileptic detection. The experiments are carried out by using time domain as well as frequency domain feature of the EEG signal. Experiments results show that epileptic detection accuracy rates as high as 99.6% with a signal input feature which is better than the results obtained by using other types signal input features.

The EEG provides an excellent tool in the diagnosis of many brain disorders, in particular, epilepsy. If a subject has exhibited definite or suspected clinical manifestation of epilepsy, routine recording of the interactical EEG are taken and closely scrutinized by an EEGer in order to detect the presence of
epileptiform activity in the form of epileptiform discharge.

Although a seemingly simple task, the scrutiny requires the EEGer to closely examine a number of channels of EEG recording of 20 minutes duration (or 20010 pages of recording) and detect the presence of epileptiform discharges manifest as spikes in one or more channels. The presence of artifacts in EEG makes the job considerably more difficult and the out come is highly dependent on the EEGer's skills. So to save time we go for automated tools for spike detection. In the current work I have tried to develop software which takes an EEG signal as input and produce a database which stores the details of each spike found in the signal.

Steps of EEG analysis

- Segmentation;
- Feature extraction;
- Classification;
- Presentation.

Adaptive segmentation based on nonlinear energy operator

The EEG is a non-stationary signal and for any automatic analysis system, it is essential that it be broken in to section that are of similar type. This involves segmenting the EEG in to quasi-stationary section and can de achieved by drawing a boundary at the instants where there is a change in the EEG pattern.

Feature extraction

The second step in the proposed EEG approach is the extraction of a feature vector that is representative of each EEG segment generated in the first step

- Artifact rejection
- Amplitude measure
- Frequency measure
- Frequency-weighted energy measure
- Multichannel feature vector
- Classification of the segmented EEG

2.1 Classification of the segmented EEG with spike and seizure detection

The third step in the EEG analysis is the classification of the segments generated in step one in to groups that are of similar types. This was first attempted by bode stein and praetorius (1977) using ward's hierarchical clustering method (ward, 1963). Other investigators have tried the probabilistic approach (bode stein et al., 1985), fuzzy cluster analysis (Amir and Gath, 1989; Krajca et al., 1991) and variant of k-means clustering algorithms9sanderson, 1980). In all work to date, classification of EEG segments has been restricted to
short duration EEG. In contrast, in the current approach, we are applying it to prolonged EEGs. To do so, we have chosen to user the k-means clustering algorithm for classification of the feature vectors corresponding to the segments.

2.2 Display of compressed EEG for clinical use

The effectiveness of the EEG is dependent on its presentation. The output of AAS-EEG is displayed in two components. The first one displays a representative segment from each of the 5 clusters. We have chosen to display the segments closest to the centriods. The second components consists of chronological sequence of EEG patterns 9time profiles0 that has been color coded to relate each segment to a particular cluster. The duration of segment on the time profile is proportional to the duration of each EEG segments in real time.

III.PREVIOUS WORK

Review on existing systems:

Some success has been achieved in automatic spike recognition of seizures primarily through rule based approach. Early rule based methods of seizures detection have been developed by gotman. These methods mostly used measures of such as amplitude, width slope and sharpness of consecutive waves.

The most widely used mimetic seizure detection method revises by gotman (1982, 1990) was evaluated by pauri et al (1992). Who concluded that the method was distinct improvement over offline visual review of EEG data from seizure monitoring while generating false positives at a rat 5.38/h.these rule based systems have become more complex in an attempt to reduce the number of false alarms of practical levels without reducing sensitivity. Recently Hao and gotman (1993) have reduced the number of false seizures detection (FSDs) to 1.26/h in a series of records selected to have a high false alarm rate. The program is trained initial FSDs for a given patients. It rejects subsequent event if they are similar to these initial FSDs for the patients.In Electroencephalogram (EEG) using Deterministic Finite Automata (DFA). It takes pre-recorded single channel EEG data file as input and finds the occurrences of Epileptic Spikes data in it. The EEG signal was recorded at 256 Hz in two minutes separate data files. A system, whose functionality was modeled with DFA, was designed. The system was tested with 10 EEG signal data files. The recognition rate of Epileptic Spike as on average was 95.68%. The result shows that the application of DFA can be useful in detection of different characteristics present in EEG signals. In this work also using pre-recorded single channel EEG data file as input and having concept of DFA function .The result shows that the application of genetic algorithm detecting spike.

IV. PROPOSED METHODOLOGY

4.1. Algorithm and Implementation

4.1.1.Preprocessing of data

The recorded EEG data is stored in a file in form of two dimensional array. The first column contains the time of recording and the second column contains the value of recorded electric potential at that time. This data was preprocessed to produce a string of '0' and '1' and was stored in an array called slope vector. The corresponding time is stored in another array called time vector. The calculation of the slope vector is done according to the algorithm as follow:

A single data point taken from the input file is represented as (x, y) where 'x' is the time and 'y' is the amplitude.

The slope of line between each pair of consecutive data points are (x1, y1) and (x2, y2) was calculated as

$$|\mathrm{Tan}\,\Theta| = |(y2-y1)/x2-x1)|$$

If computed ϴ was greater than or equal to 85^0, it shows Sharp Change and was represented by a '1', otherwise, it was taken as Nominal Change and was represented by a '0'. Then generated series of '0' and '1' is stored in array slope vector.

In the current work, spikes have been define as a combination of a certain number of Sharp Changes with certain number of allowed consecutive Nominal Changes between Sharp Changes. It has been proved that a Spike can be represented with 6 Sharp Changes and 4 consecutive Nominal Changes allowed between two Sharp Changes,

After the calculation of slope vector, Genetic Algorithm is applied on it to recognize the spikes present in it.

4.1.2 Searching for Spike

Searching phase utilizes the Genetic algorithm technique to find the spike present in the data as follows:

4.2 Selection:

The sample data (which has been used in this work as input) has been recorded with sampling frequency 256Hz and about 2 minutes data has been stored in files. So size of the Slope and Time vector will be approximately 35840.

Using random function 20 integer numbers are generated between 0 and 35840 and corresponding data from the Timeand Slope vector were recorded. It will be the initial population for the first generation.

Check for spike data (fitness function):

For each selected data, the elements of the Slope vectors starting from it in both upward and downward direction were given as input to the DFA used for recognizing the spike. The functionality of the Spike recognizer (DFASR)

[1] is explained below:

1. A string of '0' and 'I' is produced from the input signal.

2. The string is given as input to finite automata. Its scans the input string taking twenty symbols at a time and moves from one state to another depending upon the current symbol.If the automata reach to final state. One or more Spike is recognized and rest is discarded. A sub string from the input string is taken which represent the recognized spike. We repeat the procedure for the remaining part of the input string.

The finite automata for recognizing the spike is given below:-

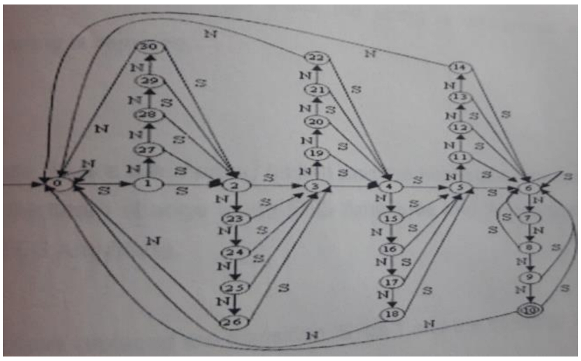

Fig: 1.5

Deterministic Finite Automata (DFA). A DFA can be defined with 5 tuples as:

M={Q, .., .., .., F}

Where,

Q- Finite non-empty set of states,

..- Finite non-empty set of input symbols,

..- Transition function, which controls the movement from one state to another depending upin the current state and current input symbol. It is defined as

………………………

Q0 – Starting state and

F – Finite non-empty set of final states

Starting with the initial state, it reads a one symbol at a time from string of input symbols and moves from current state to next state depending upon … If, at the end of the input string, the DFA reaches to a final state, the string is recognized as a valid string otherwise the string is rejected.

State 0 and state 10 are the starting (initial) and final state respectively. Transition 'S' represents the sharp changes in the EEG Amplitude and 'N' represents the nominal change in the EEG Amplitude.

In our work we have replaced the transition 'S' as 1 and the transition 'N' as 0 and If, at the end of the input string, the DFA reaches to a final state, the string is recognized as a valid string otherwise the string is rejected. Now from this we can conclude that every valid string will start with one and end up with five zeros at the end and before these five zeros there will be always a one. And also we notice for a valid string, that before the five zeros at the end we will never find five zeros consecutively.

Hence we will have two data point, one for upward direction traversal and another for downward direction traversal. The elements of the Slope data will represent a spike.

4.2.3. Production of Next Generation:

If for all the 20 selected data spike is not found then crossover and mutation were applied on the binary representation of the index of the element selected in the previous generation to produce next 20 indexes. All the elements of the new generation were checked for the spike in the same manner as it was done for the previous generation.

4.2.4. End of the process:

The process is topped as soon as we get a spike and we report the starting and ending time of it. If spike is not found then the procedure will be continued for 50 generation after that is assumed that no spike is present in the input data.

4.3. Algorithm of the proposed work:

Input: A file in which signal data is stored which consist time vector and the slope vector.

Data [ ] = Array to store slope vector which have value either '0' and '1'

X [ ] = Array to store time

Step 1 : read data from input file and store in data [ ]

Step 2: read time from input file and store in x [ ]

Step 3: initialize a count variable as 1

Step 4: selected randomly 20 time vector data and store.

Step 5: For each selected data, starting from it, sample the data from slope vector in upward and downward direction as input to the DFASR [1] to check for Spike.

Step 6: if a spike is found by the DFASR, report its start and ending time and go to end. If spike is not found do step 5 for next data in the generation.

Step 7: If for all the elements in the current generation spike is not found, generate the next population by applying crossover and mutation in the current generation.

Crossover: the selected data will be converted into binary and each data will be paired up with the other data, now in each pair of data we will interchange the bit after a randomly selected bit position. And then we apply mutation on this pairs.

SI No.	Sample data set number	Generation Number	Start time of spike	End time of Spike
1	2	1	124.445	125.527
2	2	4	63.4922	63.5625
3	2	3	109.856	109.973
4	2	1	63.4922	63.5625
5	2	7	63.4922	63.5625
6	2	2	48.1211	48.1953

Mutation: In mutation process we randomly selected a bit position and with respect to this bit position we interchange the respective bit one data with the other with in each pair.

Step 8: Repeat Step 5 and Step 6 for the new generation. If the generation number is 50, go to End.

Step 9: End.

V. RESULTS AND DISCUSSION

5.1 Result:

In the current work, spikes have been defined as a combination of a certain number of Sharp Changes with certain number of allowed consecutive Nominal Changes between Sharp Changes.

5 different pre-recorded EEG data files containing spikes were given as input to the program one by one and their output (number of generations taken to recognize the spikes using GeneticAlgorithm) were recorded. The EEG signals were recorded for two minutes separate data files to the computer hard disk after digitization of the traces at 256 Hz. The data were collected using the VisualLab-M software (ADLink Technology Inc., Taiwan). The output for all the data set (as a sample) is given in table 1 – 5

Table 1: Number of generations taken to recognize the spike and starting and end time of that spike in sample 1.

SI No.	Sample data set number	Generation Number	Start time of spike	End time of spike
1	1	6	32.1133	32.1914
2	1	2	127.887	127.957
3	1	3	120.977	121.051
4	1	1	120.977	121.051
5	1	1	32.1133	32.1914
6	1	13	127.887	127.957

Table 2: Number of generation taken to recognize the spike and starting and end time of that spike in sample 2.

Table 3: Number of generations taken to recognize the spike and starting and end time of that spike in sample 3.

SI No.	Sample data set number	Generation Number	Start time of spike	End time of spike
1	3	2	85.0898	85.168
2	3	4	35.5898	35.6641
3	3	8	59.2461	59.3945
4	3	3	44.2266	44.3398

5	3	1	85.0898	85.168
6	3	1	35.5898	35.6641

Table 4: Number of generations takento recognize the spike and starting and

SI No.	Sample data set number	Generation Number	Start time of spike	End time of spike
1	4	2	29.2031	29.2733
2	4	1	67.3672	67.4453
3	4	2	92.5938	92.6719
4	4	6	21.8359	21.9141
5	4	5	29.2031	29.2733
6	4	1	29.2031	29.2733

end time of that spike in sample 4.

Table 5: Number of generations taken to recognize the spike and starting and end time of that spike in sample 5.

SI No.	Sample data set number	Generation Number	Start time of spike	End time of spike
1	5	2	103.5	103.528
2	5	2	102.273	102.281
3	5	1	78.0039	78.589
4	5	7	88.8281	88.9063
5	5	6	7.1055	7.1797
6	5	1	7.1055	7.1797

5.2 Discussion:

The recording of EEG is done for a longer duration, since spontaneous spike activities may appear at any time in it. The manual inspection of occurrence of Spike patterns in EEG signal hence becomes a time taking and crude process. But in our method we have used genetic algorithm technique in which we first selected 20 number of data from the data files and then check these data using our fitness function, whether the data is a part of spike or not. If out these 20 data any of the data satisfies the fitness function then we record the star and end time of the spike and our objective is complete. But if all these 20 data fail to satisfy the fitness function then we generate the 2nd set of 20 data using the first selected 20 data by applying crossover and mutation and then again check among these new generated data whether any of the data satisfies the fitness function or not. This process will be repeated until any data of any generation of data satisfies the fitness function and a spike is recognized.

From the resultant data it can be easily observed that our method takes less time to recognize the spikes than the manual linear sequential process of recognizing the spikes in huge data files.

VI. CONCLUSION

Genetic Algorithm can be used successfully for search process for finding the spike pattern present in the brain signal. GA_s is mathematically modeled algorithms which emulate biological evolutionary theories to search a pattern. GA_s attempt to find the required pattern without going through an exhaustive search mechanism. Thus GA_s have an advantage of minimizing the search space and hence the search time. Here in this work a spike pattern resides in end part of the data is searched successfully in an average of 3.3 iterations, which may take huge amount of time if searched sequentially.

REFERENCES

[1] Keshri, A.K., Sinha, R.K., Hatwal, R. and Das, B.N. Epileptic Spike Recognition in Electroencephalogram using Deterministic Finite Automata. Journal of Medical Systems 33: 173 – 179, 2009

[2] Mitra S. K. and Sarbadhikari S. N. iterative function system and genetic algorithm-based EEG compression. Med. Engg. Phys. 19: 605 – 617 1997

[3] Kasabov N., Benuskova L. and Wysoski S. G. Computational Neurogenetic Modelling: Gene Networks within Neural Networks. Proc. IEEE

Intl. Joint Conference on Neural Networks, vol. 2. Pp. 1203 – 1208, 2004

[4] Cid E. A., Esparcia-Alcazar A. I. and Sharman K. Using distributed genetic programming to evolve classifiers for a brain computer interface, European Symposium of Artificial Neural Networks Bruges (Belgium), 26-28 April 2006

[5] Peterson D. A., kirbey M. J., Anderson C. W. and Thaut M. H. Feature Selection and Blind Source Separation in an EEG-Based Brain-Computer Interface Journal on Applied Signal Processing 19: 3128-3140, 2005

[6] New directions in genetic algorithm theory Gary J. Koehler, Decision and Information Sciences, University of Florida, USA

[7] Selection of Relevant Features for Classification of Movements from Single Movement Related Potentials Using a Genetic Algorithm. E. Yom-Tov, G. F. Faculty of Electrical Engieering, Technion – Israel Institute of Technology, Haifa32000, Isreal.

[8] Automatic sleep stage scoring system using genetic algorithms and neural network B.Y. Kim and K. S. Park