Proceedings - 19th International Conference - IEEE/EMBS Oct. 30 - Nov. 2, 1997 Chicago, IL. USA

Detection of Epileptic Events using Genetic Programming

Bruno Marchesi, *Member IEEE* , Álvaro Luiz Stelle, Heitor Silvério Lopes, *Member IEEE*

Abstract— **This paper presents a method using genetic programming for automatic detection of 3 Hz spike-and-slow-wave complexes, that are a characteristic of typical absences, in electroencephalogram (EEG) signals. Training features are extracted from 1s EEG frames, randomly chosen from pre-recorded files. The frames are visually classified as spike-and-slow-wave complexes (SASWC) or non-spike-and-slow-wave complexes (NSASWC). Genetic programming techniques are then applied to these data to build a program capable of recognizing such complexes.**

Keywords— **genetic programming, signal processing, EEG.**

I. Introduction

A. Epilepsy

Epilepsy is not a disease in its own right. It is, in fact a symptom caused by several different diseases while epileptic seizures are brain abnormal reactions caused by such diseases. The seizures differ depending on the extension of the affected brain area that can be involved. In general, the pathology is localized in the brain. However, important organ dysfunctions, associated with toxic-metabolic exchanges, may cause secondary encephalopathies which provoke epileptic seizures. There is also the possibility of genetic predisposition to epilepsy. It is estimated that 0.5% of the world population have epileptic seizures of some type [1]. There are three main groups of seizures, which are: 1) partial seizures, 2) generalized seizures and 3) unclassified seizures. The absence (petit mal) seizures are part of group 2.

The scalp EEG is an important resource in both diagnostic and differentiation of the various types of epilepsy, specially those associated with frequently detected cortical lesions. In general, the EEG is abnormal during the seizures (ictal period) and in between (interictal period). Spikes, isolated or associated with other waves, and sharp waves are a characteristic of such periods [2].

The ictal period of a typical absence seizure is characterized by a synchronous 3Hz spike-and-wave discharge that rarely lasts more than 10 seconds. The occurrence of such seizure is not possible if these discharges are not present, but their existence does not imply the occurrence of such a seizure.

The spike-and-wave discharge is frontally predominant but generalized and may start with a 4 Hz rate, reducing to 3 - 3.5 Hz and falling to 2.5 Hz during the final phase. Both onset and cessation of the seizure are abrupt and may be preceded and followed by normal EEG activity. The automatic detection of these complexes is essential for the analysis of long-term EEGs[3].

The authors are with the Paraná Federal Center for Technological Education (CEFET-PR / CPGEI). Av. 7 de Setembro, 3165. - 80230-901 - Curitiba - PR - Brazil. E-mails: {marchesi, stelle, hslopes}@cpgei.cefetpr.br

B. Genetic Programming (GP)

In this section, an introduction about the concepts and terminology of the GP paradigm. Its concepts are derived from the genetic algorithm (GA) [4]. The techniques employed in GP are based on the Darwinian survival and reproduction principle, which is briefly discussed next.

It is believed that natural evolution is not an oriented process for a specific purpose. Instead, there are different individuals that coexist in the same environment having survival as a common objective. Therefore, competition is naturally established. The best ones survive and have the opportunity of propagating their genetic characteristics to the next generation by means of reproduction. If the individuals are asexual, the genetic material is transmitted almost intact; otherwise, differences between individuals will emerge, but the general characteristics of the species are maintained [5].

The evolutionary process observed in nature can be applied to solve problems, by using genetic algorithm techniques which transform an object set, called population, into a new one, possibly better than the older. A performance measure, called fitness, is associated to each individual in the population and indicates how good it is for the desired solution. In this case, the better the fitness, the closer the program is to the problem solution.

This is the genetic model of learning, common to GA and GP paradigms. The main difference is in the definition of the object set. In GA, the population individuals are encoded as strings of characters, which represent chromosomes. In GP, the individuals are programs encoded as trees. To change a population of individuals into a new generation, genetic operators are employed. Such operators are: recombination (or crossover), in which the genetic material is exchanged between two individuals, and reproduction, in which an individual is integrally copied to the next population. There is also a third operator called mutation, which consists in permuting characters of an individual string, yielding a new one. This operation is usually employed in GA, but it is not so usual in GP.

In practice, each individual program is defined by a tree, instead of conventional line codes. This tree contains terminals and functions sets. Terminals may be variables, numerical values or functions that don't require an argument, such as "go up" or "go down", for instance. Functions are operators used in GP and could be arithmetic operators such as sum and subtraction. All the functions must accept all the terminals as arguments, as well as all the data types that the functions return. The choice of terminals and functions is problem dependent [6].

The following steps summarize the GP paradigm:

- create an initial population formed by a random combination of predefined functions and terminals.
- perform the following steps until a conveniently defined stop criterion is reached:

1. execute each program of the population and associate a fitness value, proportional to its performance in solving the problem.
2. create a new population of programs by means of recombination and/or reproduction.

- the best individual obtained over all generations is the GP result.

It should be noted that the stop criteria and fitness measure are defined according to the nature of the problem.

Ç. *Methods applied to pattern recognition in EEG signals*

Värri *et al.* [7] applied neural network techniques in long-term EEG signal analysis by selecting periods of special significance such as alpha waves, spike-and-wave patterns or electro-oculogram (EOG) artifacts. They obtained 92% of events detection and 13% of correctness in the discrimination of these events.

Mylonas and Comley [8] proposed a method for automatic detection of spikes using adaptive filtering and linear prediction. The method was tested only in a 10 min EEG signal, and the results obtained were: 4 false alarms, 62 detections and no lost or not detected spikes.

Kalayci and Özdamar [9] employed wavelets for preprocessing the EEG signal, before executing the automatic pattern recognition of SSWs (spike and spike and wave) and non-SSWs (any other pattern), obtaining the following results: the recognition accuracy during the tests was 2.1% lower than the training accuracy; the recognition of spike-and-wave was 50.6% of true positives for the worst case and 93.9% for the best one.

Schiff *et al.* [10] worked in the construction of an algorithm for real-time EEG processing, presenting the CSD-DWT (Critically Sampled Discrete Dyadic Wavelet Transform), based on the wavelet multiresolution property and the EEG oversampling technique. They obtained a spike detection rate of 94% in a typical EEG, with the possibility of real-time implementation.

II. Methodology

In this section, some considerations regarding the terminology used in this work will be reviewed. Next, the steps involved in the construction of the training and testing databases will be presented.

A. *Terminology*

The databases are composed by a group of features, which are numerical values calculated from a data vector. In this case, seven features containing each a 1 s duration EEG signal (called *frame*) were used. These features are those described in Hudgins *et al.* [11]: the mean absolute value (MAV), the number of zero crossings (ZC), the number of slope sign changes (SC) and the waveform length (LEN). Additionally, two other features described in Fernandez *et al.* [12] were used: the average value (AVG), the up slope (UP) and down slope (DOWN). In order to classify SASWCs and NSASWCs, the following criterion was adopted: if the frame under analysis contains 3 spikes, it is considered "characteristic" of absence discharges, and is classified as SASWC; otherwise, it is defined as NSASWC. This information is given to the GP system, associated with an additional number: 1.0 if SASWC and -1.0 if NSASWC.

B. *Training Procedure*

The procedure for the database construction, used in the training and testing steps, is the following:

- An EEG database, available on the Internet in Tampere University, Finland was used. [1]
- Two files in this database contained petit-mal discharges and were chosen to be used in this work.
- Each signal has approximately 3 min of data recorded with a sampling frequency of 200 Hz. Therefore, each channel has about 36,000 samples.
- Each file is composed by four-channels EEG records. The first two of each file were used to construct the training database and the last two for the testing database.
- Ten SASWCs and ten NSASWCs 1 s frames were randomly chosen from each of the four channels giving a total of 80 frames.
- For each one of these 80 frames, the 7 features previously mentioned were calculated, and the classification number (+1.0 or -1.0) was associated to each frame. The resulting 640 features constitute the training database.

The training database is the input raw data for the GP system. Initially, an individual program, i.e., a mathematical expression of up to seven variables (features), is randomly generated. The expression is evaluated and the result is compared to the classification number. Figure 1 shows the expected value range of an individual program.

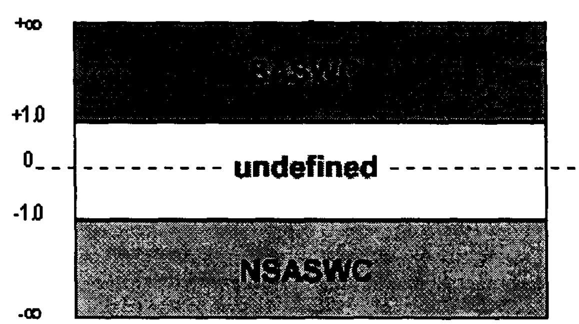

Fig. 1. Expected values for an individual program. The symbol ∞ represents an arbitrary large number.

[1] ftp://sigftp.cs.tut.fi/pub/eeg-data/ and http://www.tut.fi/

TABLE I

RESULTS OF THE DETECTION TEST.

	number of frames	correct classifications
true positive	29	36.25%
true negative	39	48.75%
false positive	1	1.25%
false negative	11	13.75%
total true	68	85%

For example, if the GP system gives 1.5 and the classification number is 1.0 (SASWC), the number of hits is incremented. The fitness criterion was defined as:

$$fitness = fitness\ cases - hits \quad (1)$$

where *fitness cases* is the number of training cases and *hits* is the match count of each individual program.

Two stop criteria were defined for the GP system. The procedure finishes if any of them is satisfied. They are:

- The number of hits of an individual program is equal to the training cases.
- A previously set maximum number of generations is reached.

When a stop criterion is reached, the performance of the best-of-all program is verified with the test database. The procedure used for creating the test database is the same used for creating the train database.

III. RESULTS

The program represented in by algebraic expression (2) was the best-of-all individual program found:

$$\begin{aligned} &2.SC.[\alpha - SC - \cos(SC)] + e^{AVG+\alpha+e^{\alpha-\cos(LEN)}} + \\ &\overline{\log}2SC + \overline{\log}2\alpha - MAV + \overline{\log}\{2SC + \\ &\cos(LEN).[MAV - AVG - SC + \\ &(AVG + 2SC).(MAV + UP)]\} \end{aligned} \quad (2)$$

where:

$$\overline{\log}x = \begin{cases} 0 & \text{if } x = 0 \\ \log|x| & \text{otherwise} \end{cases} \quad (3)$$

$$\alpha = MAV - SC + UP \quad (4)$$

Applying the program to the test database (40 SASWC and 40 NSASWC frames) the results shown in table 1 were obtained.

In order to investigate the behavior of expression (2) with an unknown EEG signal, a second experiment was conducted. It consists on sliding a 1s window throughout a 2 min single channel computing the expression (2). This experiment mimics the online use of this technique. The result is shown in figure 2. The bold curve is the computed expression, and the other one is the original EEG signal. Figure 3 shows the maximum value for the calculated expression. As can be seen, it coincides with the epileptic seizure occurrence that should be detected.

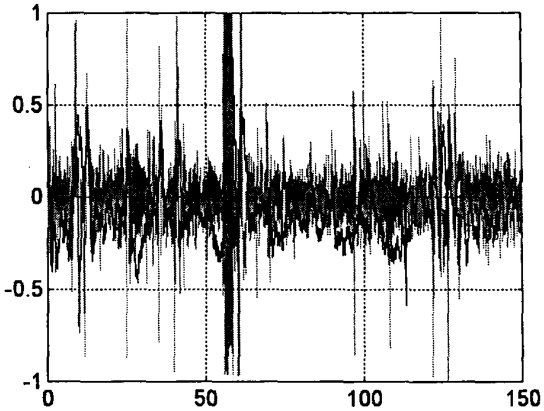

Fig. 2. Expression evaluated over the entire EEG signal of one channel. Vertical axis is relative amplitude, horizontal axis is time, in seconds.

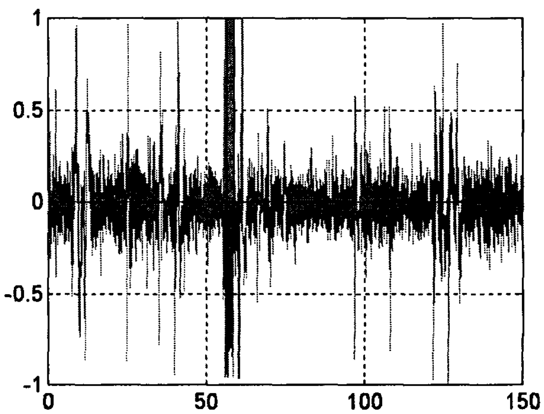

Fig. 3. Maximum value of expression 2, indicating the position where the epileptic seizure occurs. Vertical axis is relative amplitude, horizontal axis is time, in seconds.

IV. DISCUSSION AND CONCLUSION

The purpose of this work was to investigate the possibility of employing GP as an optimization method for pattern recognition programs, in the specific case of short-time EEG signals containing 3 Hz spike-and-wave complexes, which are a characteristic of absence seizures.

It has been demonstrated that satisfactory results have been obtained so far with the use of such technique. The extrapolation of the conclusions should be done with care, because the results do not point to a generalization: they are specific to the database used. However, the results strongly encourage the continuation of the experiments and studies involving this methodology.

ACKNOWLEDGMENTS

The author Bruno Marchesi would like to thank the Brazilian Government (CAPES) for the provision of financial assistance.

REFERENCES

[1] E. Nidermeyer, "Epileptic seizure disorders," in *Electroencephalography: basic principles, clinical applications, and related fields*, E. Niedermeyer and F. H. L. da Silva, Eds. Baltimore: Urban & Schwarzenberg, 1987, pp. 405-510

[2] L. A. Geddes and L. E. Baker, "Electroencepahlography," in *Principles of applied biomedical instrumentation*, L. A. Geddes and L. E. Baker, New York: J. Wiley & Sons, 1989, pp. 717-743.

[3] A. L. Stelle, *Real-time computer aided analysis of the electroencephalogram: a two-dimensional approach*, Ph.D. Thesis, The City University, London, 1991.

[4] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, Cambridge: MIT Press, 1992.

[5] J. R. Koza, *Genetic Programming*, Massachusetts: MIT Press, 1994.

[6] J. R. Koza, "Introduction to genetic programming," in *Advances in Genetic Programming*, K. E. Kinnear, Ed. Cambridge: MIT Press, 1994.

[7] A. Värri *et al.*, "Pattern recognition of the EEG in epilepsy with neural networks," in *2nd International Hans-Berger Congress: "Quantitative and Topological EEG and MEG Analysis"*, Jena, Germany, 8-11 sep, 1993.

[8] S. A. Mylonas and R. A. Comley, "Detection of epileptic spikes in the EEG using adaptive filters," in *Anais do I Fórum Nacional de Ciência e Tecnologia em Saúde, Campos do Jordão*, SP, 1992, pp. 268-271.

[9] T. Kalayci, Ö. Özdamar, "Wavelet preprocessing for automated neural network detection of EEG spikes," *IEEE Engineering in Medicine and Biology*, vol.14, n.2, pp. 160-166, 1995.

[10] S. J. Schiff, A. Aldroubi, M. Unser *et al.*, "Fast wavelet transformation of EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 91, n. 5, pp. 442-455, 1994.

[11] B. Hudgins, P. Parker and R. N. Scott, "A New strategy for multifunction myoelectric control", *IEEE Transactions on Biomedical Engineering*, vol. 40, pp. 82-94, 1993.

[12] J. J. Fernandez, K. Farry and J. B. Cheatham, "Waveform recognition using genetic programming: the myoelectric signal recognition problem", in *Proceedings of Genetic Programming Conference*, Stanford, USA, 1995.