Analysis of brain electrical signals for epileptic seizure recognition

Ashok Bhowmick

Ryerson University, Data Analytics, Big Data and Predictive Analytics
Toronto Canada

Introduction

This project aims to analyze a recent 'Electroencephalography (EEG) Data Set' with the objective of classifying epileptic seizure as compared to normal response. It is a time series data having 178 features against 1 categorical attribute through 11500 instances. Each data point represents a voltage signal corresponding to the brain activity at a point of time. Each of 178 features are one chunk out of a total of 23 created from 4097 digitized data points from one EEG record. 2 through 5 valued classes of the categorical attribute define various conditions and their subjects did not have epileptic seizures. Whereas the subjects of class 1 have seizure activity. An approach of classification and clustering will be undertaken for predictive analytics and pattern mining.

Literature Review

EEG as a technique

Electroencephalography (EEG) is a technique widely used to measure the functional activity of human brain. It is a noninvasive electrophysiological monitoring method that records the electrical activity of the brain by placing electrodes along the scalp. It measures the brain function as received at the surface of the brain in terms of voltage fluctuations resulting from ionic currents within the neurons of the brain. By that, it is a graphic display of the

difference in voltages from two sites of brain function recorded over time. Hence, EEG data is displayed as a continuous time series waveform of very small voltage signals as representative of brain function. In a normal brain the signals are of the magnitude of microvolts. The waveforms recorded are thought to reflect the activity at the surface of the brain, the cortex, however this activity is influenced by the activities from the brain structures underneath the cortex [1-2].

Instrument and Data Recording

The recording of EEG data is done by placing small metal plates as electrodes at specially designated places on the scalp following an international 10/20 system. For adults, there are 21 measuring electrodes and 2 reference electrodes (total 23). Each site is labelled by the combination of a letter and a number. The letter refers to the part of the brain, like F - frontal lobe, T - temporal lobe, C - central lobe, P – posterior lobe etc. whereas any even numbers designate the right side of the head as opposed to any odd number designating the left part of the brain, as shown in the figure below. A1 and A2 are the two reference electrodes. Voltage signals are measured with reference to the corresponding reference electrodes. Electrodes can be placed individually or by a cap that has electrodes attached onto it [3-4].

EEG machines use differential amplifier that measures the difference potentials as inputs from individual electrodes. In common reference derivation, each amplifier measures the difference of voltages between a scalp and a reference as shown in Fig. 1. The manner, in which pairs of electrodes are connected to each amplifier is called a montage. Each montage uses one of the three standard recording derivations, e.g. common reference, average reference and

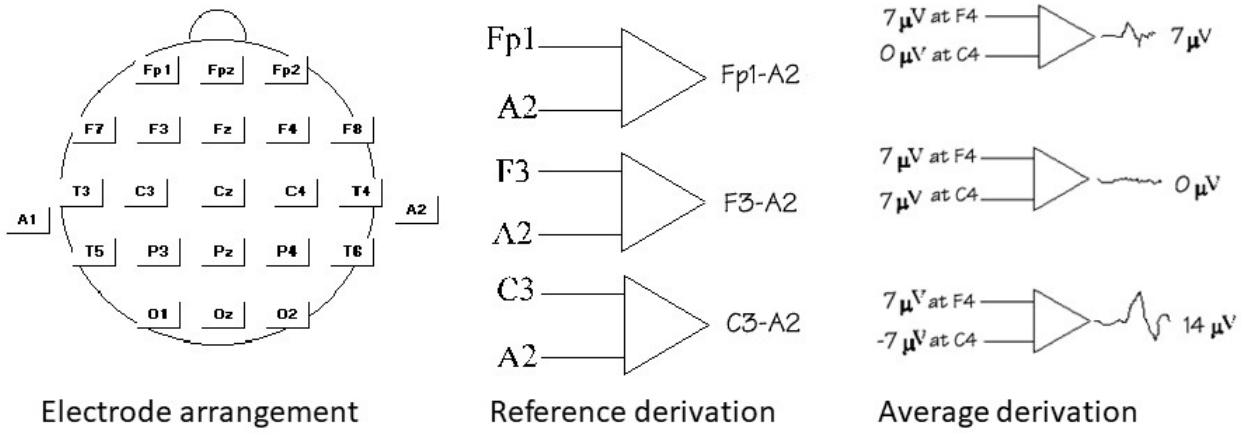


Figure 1. Electrode arrangement and reference derivation examples to record an EEG data

bipolar. The magnitude of the signal is the difference potential of which the sign depends on the sign of the higher potential. Therefore, the values at a given time can either be positive or negative voltage. This makes the correlation of the output of an individual electrode with a neurophysiological phenomenon complex. In an analog machine, the signal is recorded as a continuous pattern and in a group formed by all electrode responses, they are referred as the brain wave. Modern machines digitize the signal using an ADC which can be stored, displayed and manipulated using a computer. The rate at which the signal is digitized is called the sampling rate which typically ranges between 50 – 300 Hz [5].

What EEG data represents

Electrical signals are created when charges move within the central nervous system which originates from cerebral potential based upon the electrophysiological properties of the nervous system at the time of measurement. When neurons are activated, local current is produced due to synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex. Differences of electrical potentials are caused due to creation of electric dipoles between soma or the body of neuron and apical dendrites or the neural branches. A current is thereby generated due to the transport of Na^{+}, K^{+}, Ca^{+2}, and Cl^{-} ions that are pumped through neuron membranes following the gradient of membrane potential [1, 6]. The signal forms wave shape that are commonly sinusoidal where peak to peak magnitude vary between 0.5 to 100 microvolts (for healthy brain) and the frequency can range from 0 to half of the sampling frequency. The signals are normally collected instructing the subjects to keep their eyes closed and be in relaxed situation. The waves are categorized into four basic groups viz. 1) Delta: 0.5 to 4 Hz, 2) Theta: 4 to 8 Hz, 3) Alfa: 8 to 13 Hz and 4) Beta: all greater than 13 Hz, as shown in the Fig. 2 [4, 7]

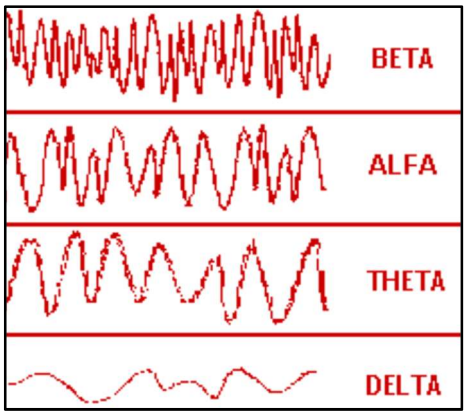


Figure 2. Types of Brain waves

Interpretation of EEG data

Interpretation of EEG signal for diagnostics is an involved task. Similar pattern can be observed for many different reasons. Widely varied patterns might result from the same symptom as well because the signals depend on many different factors and are unique to

individuals. This enables to distinguish persons by only recording their brain activity. For example, rational subjects demonstrate higher activity on their frontal left hemisphere as opposed to the intuitive subjects having strong activity on frontal right hemisphere. However, there is no thumb rule that can be formulated to be applied universally. Alpha activity is induced by closing the eyes and by relaxation and abolished by eye opening or alerting by any mechanism like, thinking or calculating. Subjects are found to be remarkably sensitive to eye closing condition, i.e. when the eyes are closed the wave pattern significantly shifts from beta to alpha waves. However, the precise origin of the alpha rhythm is still not known in detail [1, 8]. That way, EEG is sensitive to a continuum of states ranging from stress state, alertness to resting state, hypnosis, and sleep. During normal state of wakefulness with open eyes, beta waves are dominant. In a state of relaxation or drowsiness, alpha activity rises. Sleep is generally divided into two broad types: nonrapid eye movement sleep (NREM) and rapid eye movement (REM) sleep that occur in alternating cycles. Various regions of the brain do not emit the same brain wave frequency simultaneously. An EEG signal between electrodes placed on the scalp consists of many waves with different characteristics. A large amount of data received from even one single EEG recording presents a difficulty for interpretation. To deal with this complexity, it is always difficult and critical to organize a digitized EEG data for critical analysis and interpretation.

An abnormal pattern in EEG data may support diagnosis, indicate cerebral dysfunction or can be due to something different than what the study has been conducted for [1-2, 7]. Clinical applications of EEG span a very broad range of diagnostics conditions [4]. Such as,

(1) monitor alertness, coma and brain death;

(2) locate areas of damage following head injury, stroke, tumour, etc.;

(3) test afferent pathways (by evoked potentials);

(4) monitor cognitive engagement (alpha rhythm);

(5) produce biofeedback situations, alpha, etc.;

(6) control anaesthesia depth ("servo anaesthesia");

(7) investigate epilepsy and locate seizure origin;

(8) test epilepsy drug effects;

(9) assist in experimental cortical excision of epileptic focus;

(10) monitor human and animal brain development;

(11) test drugs for convulsive effects;

(12) investigate sleep disorder and physiology.

Epilepsy and EEG

This work will focus only on epileptic abnormality. Epilepsy is a chronic disorder and is distinguished by recurrent unprovoked seizures [9-10]. A person is diagnosed with epilepsy if they have two unprovoked seizures (or one unprovoked seizure with the likelihood of more) that were not caused by some known and reversible medical condition like alcohol withdrawal or extremely low blood sugar. The seizures in epilepsy may be related to a brain injury or a family tendency, but often the cause is completely unknown. The word "epilepsy" does not indicate anything about the cause of the person's seizures or their severity. Many people with epilepsy have more than one type of seizure and may have other symptoms of neurological problems as well. Sometimes, group of people having epilepsy are found to have similar clinical history, family history, outlook and EEG data. In these situations, their condition is referred as specific epilepsy syndrome. In summary, epilepsy can be characterized as follows,

1. It is the fourth most common neurological disorder and affects people of all ages.
2. It means the same as "seizure disorders"
3. Its symptom is unpredictable seizure and can also cause other health problems
4. Its control varies from person-to-person
5. There is considerable amount of public misunderstandings on epilepsy
6. Proper awareness and public educating program is still lacking.

An unusual EEG pattern can be indicative of an epileptic seizure and can act as a fingerprint to reveal which part of the brain is affected through an identification of the electrode that produced the unusual response as compared to a data from normal brain. However, in some cases the brain shows normal behavior when there is no seizure. In other cases, unusual activity is observed all the time which may be due to several reasons like tumor growth etc. Recording of an *ictal* discharge i.e. EEG pattern when the seizure occurs is often very time consuming and laborious in an outpatient setting. *Interictal* discharges are therefore recorded using portable devices in home settings. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) describes interictal discharges as a subcategory defined as "distinctive waves or complexes, distinguished from background activity". Interictal epileptiform has spiky morphology, however all other criteria must be verified empirically for diagnosis [11-12]. Analysis of an interictal discharge can therefore reveal several information

regarding the seizure. Sufficient number of case studies can therefore be grouped with respect to external and behavioral parameters to generate a generalized model that might help alleviate the diagnosis of unknown cases quicker and contribute in a better seizure prevention strategy.

Dataset

The data set for this project is taken from UCI Machine Learning repository. It is a multivariate timeseries having 178 features against 1 categorical attribute through 11500 instances. Each data point represents a voltage signal corresponding to the brain activity at a point of time.

The original data is composed for the study as five sets denoted by A, B, C, D, E each containing 100 single channel EEG segments of 23.6 seconds duration. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme. Volunteers were relaxed in an awake state with eyes open which is denoted as data - A and eyes closed, denoted as data - B, respectively. Sets C, D, and E originated from an EEG archive of presurgical diagnosis in Germany. The archive data is selected from five patients all of whom achieved complete seizure control after section and their epileptic zone were correctly diagnosed. The data-D was recorded from within the epileptic zone whereas data-C was collected from the hippocampal formation of the opposite hemisphere of the brain [13].

The data that is being dealt with here has been reorganized from this original data for machine learning studies as posted in UCI Machine Learning repository [14]. Each time series record A—E has been digitized to 4097 data points out of which 23 chunks have been created distributing into 178 time-features of recorded raw-voltage signals. 5 sets of 100 EEG segments thus creates (5×100×23) 11500 instances. The grouping on 23 chunks corresponds to the fact that there are 23 electrodes used in an EEG recording. A class attribute or output as 'y' is generated as follows:

Class 5 : Eyes open condition from healthy volunteer i.e. data-A, EEG is collected keeping the eyes open

Class 4 : Eyes closed condition from healthy volunteer i.e. data-B, EEG is collected keeping the eyes closed

Class 3 : Tumor is identified and located but the data is collected at the nonepileptic healthy part of the brain (archive data)

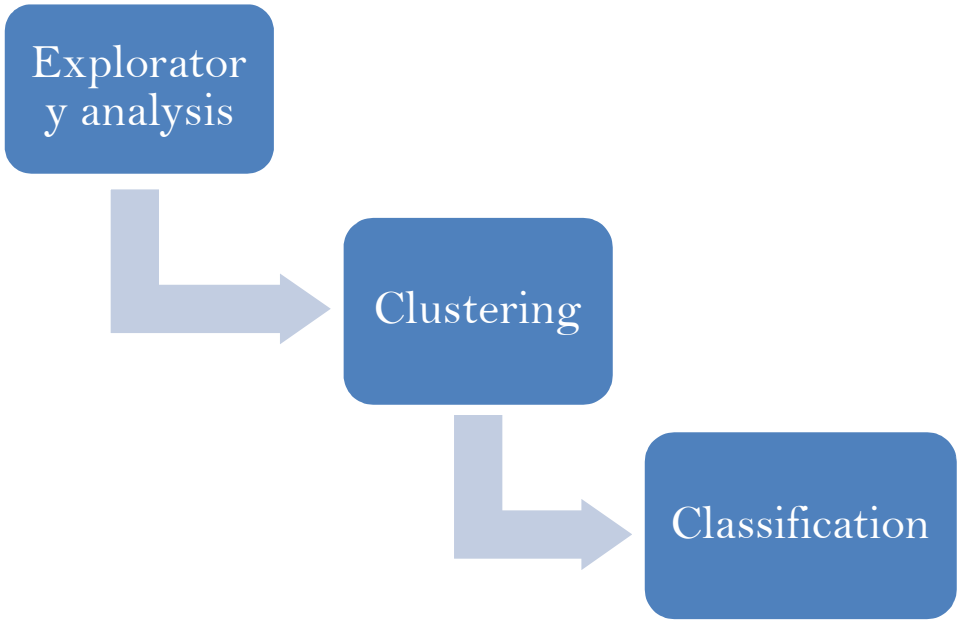
Class 2 : Tumor is identified but data is collected from the epileptic zone of the brain

Class 1 : Recording during seizure activity i.e. the *ictal* data.

As described, the data is a time stamp of recorded voltage in microvolt unit. There are no missing value in the data set.

Approach

From the structure of the data set and its class description, it appears to a machine learning problem of classification and clustering. Subjects of class 2—4 are nonepileptic whereas only the subjects of class-1 are epileptic. Most likely treatment of a binary classification would be epileptic versus nonepileptic where 2—4 classes are designated as value 0 and class-1 as value 1. However, clustering might reveal interesting information as well.



Step 1: Exploratory Analysis

Exploratory analysis of the data has been done in Python. The source code is uploaded in Github repository. All analysis will be eventually compiled onto the final report along with derived feature analysis plots for predictive analytics. Preliminary analysis shows the following distinct features:

1. There are no missing data
2. Outliers are there in some attributes, but fall within the distribution pattern and not suggestive of removing
3. The distribution over timestamps (column) is very uniform, however not normal (or Gaussian). It could be close to a Lorenzian type (to be explored). The values ranges between ±500 μV.
4. Distribution over instance show the frequency pattern varying widely between very small value (±50 μV) to as large as ±1 mV. This needs to be explored against the class values because an *ictal* data always exhibits larger magnitude.
5. There are 2300 instances for each class 1—5 and the distribution is even. The sum of class-1 instances varies over small margin from one timestamp to the other,
6. The .md and jupyter file can be found in my Github repository.

Link: https://github.com/ashokbhow/Capstone-Project

Step 2: Clustering

An approach to clustering analysis using k-means algorithm will be worked out to see how a regrouping correspond to the given class description (R/Python)

Step 3: Classification

Weka M5P and Perceptron classification have been performed on the whole dataset on a 66% / 33% train and test split. This result will be explained in final report in comparison to other methods.

Step 4: Entropy analysis

Entropy analysis on class-1 epileptic instances may be a good approach to reveal the degree of regularity in ictal data which may be symptomatic to the type of epilepsy. Details will be worked out and provided in final report [15].

References / Literature

[1] William O. Tatum, Handbook of EEG Interpretation – Normal EEG, (2008), Demos Medical Publishing LLC, ISBN-13: 978-1-933864-11-2 (pbk: alk paper)

[2] Niedermeyer E., da Silva F. L., (2004), Electroencephalography: Basic Principles, Clinical Applications and Related Fields, Lippincott Williams & Wilkins, ISBN 0-7817-5126-8.

[3] https://medlineplus.gov/ency/article/003931.htm

[4] M. Teplan, Measurement Science Review, 2, (2002), Section 2

[5] E. J. Smith, in http://www.ebme.co.uk/articles/clinical-engineering/56-introduction-to-eeg

[6] H. L. Atwood, W. A. MacKay. 1989. Essentials of neurophysiology, B.C. Decker, Hamilton, Canada.

[7] Selim R. Benbadis, Handbook of EEG Interpretation – Abnormal Nonepileptiform EEG EEG, (2008), Chapter 2, Demos Medical Publishing LLC, ISBN-13: 978-1-933864-11-2 (pbk: alk paper)

[8] R.D. Bickford. 1987. Electroencephalography. In: Adelman G. ed. Encyclopedia of Neuroscience, Birkhauser, Cambridge (USA), 371-373.

[9] William O. Tatum and Selim R. Benbadis, Handbook of EEG Interpretation – Abnormal Nonepileptiform EEG EEG, (2008), Chapter 3, Demos Medical Publishing LLC, ISBN-13: 978-1-933864-11-2 (pbk: alk paper)

[10] https://www.epilepsy.com/learn/about-epilepsy-basics/what-epilepsy

[11] https://emedicine.medscape.com/article/1138880-overview#a1

[12] Bergamini, M. Dondey, D. W. Klass, M. Lennox-Buchthal and I. Petersén, Electroencephalogr. Clin. Neurophysiol. 37(5), 1974, p. 538

[13] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger, Phys. Rev. E, 64 (2001) p.61907-15

[14] Qiuyi Wu and Ernest Fokoue, School of Mathematical Sciences, Rochester Institute of Technology, UCI Machine Learning Repository, Date uploaded: May 24 2017

[15] N. Kannathal, Min Lim Choo, U. Rajendra Acharya, P. K. Sadasivan, Computer Methods and Programs in Biomedicine, 80 (2005) p.187