Applying Data Mining for Epileptic Seizure Detection

Ying-Fang Lai[1] and Hsiu-Sen Chiang[2*]
[1]Department of Industrial Education, National Taiwan Normal University
162, Heping East Road Sec 1, Taipei, Taiwan
[2]Department of Information Management,
National Taichung University of Science and Technology,
129, Sec 3, Sanmin Rd, Taichung City 404, Taiwan ROC
[1]maymayparadise@dwvs.cy.edu.tw; [2*]hschiang@nutc.edu.tw

Abstract: - Epilepsy is a common neurological disorder of the brain to produce sudden illness. The incidence is surpassed only the cerebrovascular. Characterized by measuring brain waves can objectively detect whether subjects are suffering from epilepsy. We analyze physiological parameters of electroencephalogram and retrieve a plurality of sub-band through discrete wavelet transform. Datamining methods including multilayer perceptron neural network, support vector machine and Bayes nets were develops diagnostic modes for epilepsy detection. This result finds Bayes nets had a better performance than other datamining methods.

Key-Words: - Electroencephalograph, Epilepsy, Discrete Wavelet, Data Mining

1 Introduction

Epilepsy is a common neurological disorder of the brain to produce sudden illness. According to statistics, the majority of epilepsy happens on the age under 19, while about 80% happens on the age under 10. Epilepsy is a cell inside brain caused by abnormal discharge and the symptoms include delirium, body twitching, trismus, foaming at the mouth, etc.

In the clinical detection and diagnosis of epilepsy, EEG signal is as the main tool. Epilepsy is recurrent and often has abnormal brain discharges which can damage brain cells, thus, the treatment will be more difficult. Drug treatment is the main way to treat epilepsy currently, however, there is a certain percentage of patients been using a variety of drugs still unable to be controlled. Non-drug adjunctive therapy even is required such as epilepsy surgery.

Purpose of this study was through discrete wavelet transform to analyze physiological parameters of EEG and retrieve a plurality of sub-band. Adopted many classifications including support vector machine (SVM), neural network (NN) and Bayes nets (BNs) to compare the diversity data to see which is more suitable for classification of epilepsy electroencephalogram (EEG) data, develops diagnostic mode, and then save a lot of time and cost.

2 Related Works

2.1 Epilepsy

Epilepsy is the name of brain disorder from the cause of innate or acquired characterized by the excessive discharge of brain cells and repeated seizures. Multiple clinical symptoms usually accompany convulsions or unconsciousness and others. Epilepsy is a complex disorder, however, very small proportion epilepsy cases, less than 1/3, show genetic transmission. Except for the sporadic cases, anticonvulsant drugs are used to and acted on the brain to reduce the frequency and severity of seizures. Epilepsy can be divided into generalized seizures and partial seizures:

(1)Partial seizures (remains in a limited area of the brain)

- Simple partial seizures: it do not affecting awareness or memory
- Complex partial seizures: affecting awareness or memory of events.
- Partial seizure evolving to secondarily generalized seizure

(2) Generalized seizures

- Absence seizure
- Myoclonic seizure
- Tonic seizure
- Tonic-clonic seizure
- Atonic seizure
- Clonic seizures

2.2 Electroencephalography

Electrical activity of the human brain happens all the time. The currents on the cerebral cortex are composed of cells with the potential difference between the other cell populations generated. Those activities were recorded from the scalp surface after being picked up by metal electrodes. This is called brain wave, also EEG. EEG has several strong points as a tool for exploring brain activity.

EEG, with high temporal resolution, can directly and immediately recording operation of the brain. It is commonly used for the non-invasive clinical diagnosis and longitudinal tracking of brain disease. Most of the cerebral signal observed in the scalp EEG falls in the range of 1~20 Hz (activity below or above this range is likely to be artefactual, under standard clinical recording techniques). Waveforms from low to high are subdivided into bandwidths known as delta(0.5~4Hz), theta(4~7Hz), alpha(8~15Hz), and beta(16~31Hz) to signify the majority of the EEG used in clinical practice (Tatum, 2014).

3 Methodology

The brainwaves are divided into various sub-bands based on the discrete wavelet transform. The characteristics of brainwaves are extracted from various sub-bands. Also, different classification methods are adopted to detect epilepsy and their performances are compared with each other.

3.1 Discrete wavelet transform

Discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. It captures both frequency and location information and mainly takes the raw data into a data consistency and a high degree of variability in the data. The DWT of a signal is calculated by passing it through a series of filters. The number of columns generated by the low pass filter preserves the original number of columns of data consistency. While the high frequency filter will remain high variability in the number of columns of data (Daubechies, 1990).

The structure of discrete wavelet transform is shown as Fig. 1. $x[n]$ is the input signal and N shows the length. $g[n]$ is the low pass filter which passed through with impulse response and then output $L[n]$. $h[n]$ represents high pass filter and then output $H[n]$. $\downarrow Q$ is downsampling filter. “a” is the “a” level of in the wavelet transform structure.

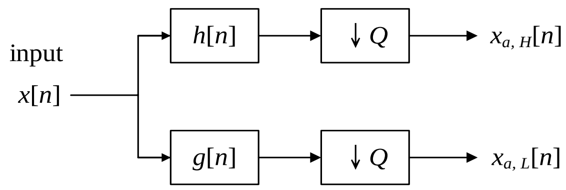

Fig. 1 The structure of Discrete Wavelet Transform

3.2 Support vector machine

Support Vector Machine (SVM) is supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. A SVM training algorithm builds a model that assigns new examples into one category or the other. A good separation in SVM is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (Bastı et al., 2015).

3.3 Neural network

In an artificial neural network, simple artificial nodes, known as "neurons", "neurodes", "processing elements" or "units", are connected together to form a network which mimics a biological neural network. Neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, thus it is applied into carious field.

3.4 Bayes net

Bayes Net (BN) which represents a set of random variables and their conditional dependencies. Creating various features of incidence and then calculating conditions probability of any set of variables in the network. Bayes network also has the ability of collocation and prediction. Bayes network also has the ability of collocation and prediction.

4 Experiment and Result

4.1 Samples

The dataset in the study are adopted from five different categories of patients. These data were accessible from information set (Andrzejak et al., 2001). EEG signals were often used to detect epilepsy in past researches (Güler and Ubeyli, 2007, Tzallas et al., 2009, Ubeyli and Guler, 2007 and Ubeyli, 2008). Category 1 and 2 represent the brain wave for healthy subjects, eyes open and eyes closed, respectively. Category 3 shows the brain waves measured from surrounding brain hippocampal formation in epileptic patients. Category 4 is epilepsy seizure EEG concentrated on the recording between the hippocampal formations in patients. Category 5 is concentrated on the hippocampus from the structure confirmed by the clinician. Each category contains 100 data, 500 as a total. Every 23.6 seconds is divided into a sampling frequency in the range of 0.53-40 Hz, and the sampling rate is 173.61Hz, with a resolution of 12bits. The study divides category 1 and 2 into as a Set A (Non-epileptic subjects) and category 3~5 is as Set B (epileptic subjects).

4.2 Data processing

Conversing 500 data through discrete wavelet transform function built-in MATLAB R2012a, db2 is a wave have a smoothing effect (Güler et al., 2005). First, after the first discrete wavelet transformation, through a high-pass filter and a low-pass filter will produce sub-bands D1and sub-bands A1. Then use the sub-band A1 proceed the second discrete wavelet transformation, generate sub-band D2 and sub-band A2. Taking A2 to use a third discrete wavelet transformation, and so on. There are total 4 times of wavelet conversion. The Fig. 2 shows all of sub-bands D1 ~ D4 and A4.

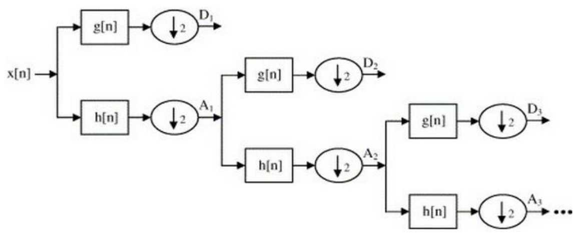

Fig. 2 Discrete wavelet transform process

4.3 Features extraction

According to previous literature review, after the discrete wavelet transform of each sub-band, the value of maximum, minimum, mean, and standard deviation (SD) are the EEG characteristics of epilepsy classification. Set A is as for the non-epileptic subjects and Set B is for epileptic subjects are shown in Table 1.

4.4 Evaluation of Epilepsy seizure detection

This section showed the comparison of the classification methods in the diagnosis of epilepsy. These following indicators were used to evaluate the effectiveness of these methods includes true positives rate (TPR), true negatives rate (TNR), false negatives rate (FNR) and false positives rate (FPR). The precision rate is defined as TP/(TP+FP) and recall rate is TP/(TP+FN). Accuracy is defined as (TP +TN) / (TP +FP + FN + TN). The *F*-measure combines precision TP/(TP + FP) and recall TP/(TP + FN) on the prediction of the positive class (*F*-measure = 2 × precision × recall / precision + recall). AUC is the area under the receiver operating characteristic curve calculated with MedCalc.

Different datamining methods are used to develop detection model including: A multilayer perceptron neural network (NN), support vector machine (SVM), and Bayes net (BN). The polynomial kernel is a kernel function commonly used with SVM. The dataset is divided to 80% for training model and 20% for testing model. The results of D1~D4 and A4 sub-bands were shown in Table 2~6. From the assessment in Table 2~6， Bayes net has the highest performance in all measure indexes in all of sub-bands. The MLP-NN has slightly higher accuracy rate than SVM in sub-bands D1 and A4. Also, SVM is slightly higher in D2 and D4 than MLP-NN. All methods in the D4 sub-bands are lower than 70% and have poor performance.

Table 1 Features description of various Sub-bands

Dataset	*Extracted features*	*Sub-bands*				
		D_1	**D_2**	**D_3**	**D_4**	**A_4**
Set A	Maximum	125.54	233.66	652.19	903.58	787.93
	Minimum	-125.85	-289.51	-550.85	-852.35	-786.9
	Mean	-0.003	-0.004	0.031	-0.808	-37.572
	SD	2.994	10.379	36.96	48.755	29.684
Set B	Maximum	779.05	1548.7	3339.7	3789.2	6094.8
	Minimum	-863.56	-1847.7	-3710	-4376.5	-5339.8
	Mean	-0.005	-0.02	-0.06	-0.03617	-26.470
	Standard deviation	21.54	89.05	229.356	342.812	350.364

Table 2 Performance measure of different approach in Sub-band D1

Methods	TPR	FPR	Precise	Recall	AUC	F-measure	G mean	Accuracy
MLP-NN	0.73	0.216	0.779	0.73	0.852	0.733	0.805	73.00%
SVM	0.62	0.62	0.384	0.62	0.500	0.475	0.634	62.00%
BN	0.82	0.161	0.835	0.82	0.887	0.822	0.862	82.00%

Table 3 Performance measure of different approach in Sub-band D2

Methods	TPR	FPR	Precise	Recall	AUC	F-measure	G mean	Accuracy
MLP-NN	0.57	0.376	0.628	0.57	0.567	0.573	0.645	57.00%
SVM	0.62	0.62	0.384	0.62	0.5	0.475	0.634	62.00%
BN	0.86	0.106	0.881	0.86	0.938	0.862	0.909	86.00%

Table 4 Performance measure of different approach in Sub-band D3

Methods	TPR	FPR	Precise	Recall	AUC	F-measure	G mean	Accuracy
MLP-NN	0.62	0.62	0.384	0.62	0.551	0.475	0.634	62.00%
SVM	0.62	0.62	0.384	0.62	0.5	0.475	0.634	62.00%
BN	0.79	0.149	0.841	0.79	0.875	0.793	0.868	79.00%

Table 5 Performance measure of different approach in Sub-band D4

Methods	TPR	FPR	Precise	Recall	AUC	F-measure	G mean	Accuracy
NN	0.56	0.647	0.418	0.56	0.61	0.459	0.578	56.00%
SVM	0.62	0.62	0.384	0.62	0.5	0.475	0.634	62.00%
BN	0.65	0.215	0.818	0.65	0.718	0.636	0.806	65.00%

Table 6 Performance measure of different approach in Sub-band A4

Methods	TPR	FPR	Precise	Recall	AUC	F-measure	G mean	Accuracy
NN	0.77	0.172	0.82	0.77	0.817	0.773	0.847	77.00
SVM	0.71	0.198	0.804	0.71	0.756	0.709	0.819	71.00
BN	0.87	0.11	0.882	0.87	0.904	0.872	0.906	87.00

As the mention above, Bayes net is the best way for epileptic seizure detection, especially in the sub-band A4. Therefore, comparing with other methods, the Bayes net can handle large data sets in a short time. MLP-NN and SVM in sub-bands D1~D4 have poor performance for epilepsy detection. So, they are not suitable for using the classification of epileptic EEG data. A sub-band A4 has the highest classification results in accuracy rate and suitable to be used to detect epilepsy based on EEG data.

5 Conclusion

In order to save time and cost, this research proposed that the usage of a Bayes net in the sub-band A4 with the highest accuracy (87%). Epilepsy seizure detection based on EEG may not only encounter more problems of signal processing and analysis, but also cause false positives because of the high complexity of the disease. Application of datamining methods for epilepsy seizure detection, Can assist physicians quickly understand the condition and as a basis to provide advice on clinical diagnosis.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology of Taiwan for Grants MOST 104-2410-H-025-023.

References:

[1] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, Indications of nonlinear deterministic and finite-dimensional

structures in time series of brain electrical activity: Dependence on recording region and brain state, *Physical Review E*, Vol. 64, 2001.

[2] E. Bastı, C. Kuzey and D. Delen, Analyzing initial public offerings' short-term performance using decision trees and SVMs, *Decision Support Systems*, Vol. 73, 2015, pp.15-27.

[3] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Transactions on Information Theory*, Vol.36, No.5, 1990, pp.961-1005.

[4] J. Engel, *Seizure and Epilepsy*, Philadelphia, PA: Davis, 1989.

[5] N. F. Güler, E. D. Ubeyli and İ. Güler, Recurrent neural networks employing Lyapunov exponents for EEG signals classification, *Expert Systems with Applications*, Vol. 29, No.3, 2005, pp.506-514.

[6] I. Güler and E. D. Ubeyli, Multiclass support vector machines for EEG-signals classification, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 11, No.2 pp. 117–126, 2007.

[7] E. D. Ubeyli and İ. Güler, Features extracted by eigenvector methods for detecting variability of EEG signals, *Pattern Recognition Letters*, Vol. 28, 2007, pp.592-603.

[8] E. D. Ubeyli, Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, *Computers in Biology and Medicine*, Vol. 38, No.1, 2008, pp. 14-22.

[9] E. D. Ubeyli, Probabilistic neural networks combined with wavelet coefficients for analysis of electroencephalogram signals, *Expert Systems*, Vol. 26, No.2, 2009, pp.147-159.

[10] E. D. Ubeyli, Decision support systems for time-varying biomedical signals: EEG signals classification, *Expert Systems with Applications*, Vol. 36, 2009, pp.2275-2284.

[11] W. O. Tatum, Ellen R. Grass Lecture: Extraordinary EEG, *Neurodiagnostic Journal*, Vol. 54, No.1, 2014, pp.3-21.

[12] A. T. Tzallas, M. G. Tsipouras and D. I. Fotiadis, Epileptic seizure detection in EEGs using time–frequency analysis, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 13, No.5, 2009, pp.703-710.