Metric Classification of Traumatic Brain Injury Epileptiform Activity from Electroencephalography Data

Yu.V. Obukhov[1], K.Yu. Obukhov[2], S.A. Nikitov[1]

[1]Kotel'nikov Institute of Radioengineering and Electronics RAS, Mokhovaya 11-7, Moscow, Russia, 125009

[2]Moscow Insitute of Physics and Technology (State University), Institutskiy per. 9, Dolgoprudny, Moscow Region, Russia, 141701

Abstract. Prediction algorithm of Epilepsy Seizures and Sleep Spindles in electroencephalography (EEG) data is studied in this article. EEG data was measured in rats with Post-Traumatic Epilepsy (PTE) before and after Traumatic Brain Injury (TBI). Experts manually partitioned records into two classes: one, which refers to epileptic activity - Epilepsy Seizures, and second class, which refers to normal behavior of rats - Sleep Spindles (SS). Proposed algorithm was trained and tested on the collected data, which contained EEG features, previously extracted by detection algorithm. Feature importance was evaluated, and logistic regression model was built. Cross validation results were 79% Area Under Curve (AUC) for the best model.

Keywords: Traumatic brain injury, EEG, Wavelet, Spectrogram, Ridges, Event detection, Epileptiform seizures, Metric classification, Logistic regression.

1. Introduction

The wide known fact is the connection of epilepsy and head injury. According to statistics, significant number of epilepsy cases happen after Traumatic Brain Injury (TBI). This kind of epilepsy is called Post-Traumatic Epilepsy (PTE). As the Epilepsy Seizures (ES) are one of the strongest features of PTE risk, it is suggested that these seizures are ideal target for early disease diagnostics. Among other phenomena in EEG, for instance, Sleep Spindles (SS) refer to normal behavior in brain activity. Thus the recognition of events in EEG, like ES and SS poses significant problem to biomedical scientific community.

2. Experiment and Data Collection

In order to create a recognition algorithm, one should collect the training data of EEG activity. Thus, the experiment was set up on rats with PTE. EEG records were measured before and after TBI in order to receive unbiased data. After the EEG data was collected, Experts performed a manual mark up of signals into two classes of ES and SS.

Raw EEG data was processed by continuous wavelet transform in order to receive wavelet spectrogram of Power Spectrum Density (PSD), Frequency and Time. EEG records were measured in 4 Channels. List of extracted features in EEG is provided in a Table 1.

As a result, the dataset of 365 observations was made, 198 of observations were labeled as ES and 167 as SS. Along with 10 features, a target variable was added. This variable indicated class of a record according to mark up: **1** in case of epilepsy (ES) and **0** in case of normal (SS) observation.

Table 1. List of EEG Features and their AUC values

Feature	Description	AUC
f_{min}	Minimal Frequency of a record	0.52
f_{max}	Maximal Frequency of a record	0.51
Δf	f_{max} - f_{min}	0.51
P_{min}	Minimal PSD of a record	0.72
P_{max}	Maximal PSD of a record	0.53
ΔP	P_{max} - P_{min}	0.5
mean(f)	Average Frequency of a record	0.61
std(f)	Frequency Standard Deviation of a record	0.64
std(f)/mean(f)	Standard Deviation to Mean ratio of Frequency	0.67
t(P_{max})	Relative Time when maximal PSD was observed	0.65

3. Model Creation and Validation

3.1. Feature Importance

Logistic Regression model was trained on a dataset. However, only important features were used as predictors. In order to estimate feature importance, Receiver Operating Characteristics (ROC) were built and Area Under Curve (AUC) was measured. Table 1 also provides AUC results for the features. AUC 0.5 indicates that there is no dependency between feature and target variable. As a result, only important features with AUC>0.6 were used in a model.

3.2. Model Training and Testing

Dataset was randomly divided into train and test parts in 70/30 proportion. Coefficients evaluation was performed on a train subset, and prediction was made on test subset. As a result, AUC was measured on a validated data. This training and testing procedure was made 1000 times in order to reach statistically stable results.

4. Results

After 1000 models were created, AUC for each model was estimated on a test subset. As a result, distribution of AUC was analyzed. Table 2 shows AUC statistics, and it can be seen that median AUC is nearly 80%. Average coefficients of predictors are provided in a Table 3.

Table 2. Summary of model's Area Under Curve values

	Min.	1st Quartile	Median	Mean	3rd Quartile	Max
AUC	0.6425	0.7692	0.796	0.795	0.8217	0.8959

Table 3. Average features coeffiecients in Logistic Regression models

Intercept	P_{min}	mean(f)	std(f)	std(f)/mean(f)	t(P_{max})
-1.8203	-3.682e-05	0.5165	-1.5043	1.9391	0.6578

Based on these coefficients, probability of ES for a particular record can be estimated using formula:

$$P = \frac{1}{1 + e^{-z}}$$

$$z = k_0 + \sum_{n=1}^{5} k_n F_n$$

In a formula above F_n indicates feature values and k_n refers to their coefficients from Table 3 (k_0 refers to Intercept).

5. Conclusion

This article describes methods of recognition of epileptic signals in brain activity. EEG records of laboratory rats with PTE were measured before and after TBI. Experts in neurobiology performed a mark up of signals into ES and SS in order to make a dataset for training the model. Features were extracted from EEG via wavelet transform and time-frequency analysis. On a resulting dataset, logistic regression model was built with binary target variable of epilepsy. Although, only important factors were considered in a model. Models were cross-validated on a subset of data and performance was measured in AUC.

Proposed algorithm of ES and SS recognition shows good performance results with average of 80% accuracy of prediction. Further, this approach can be applied to identification of epileptic activities in a long term EEG records. For instance, long EEG records can be automatically processed, and parts with high risk of epilepsy could be identified with proposed model. Thus, dynamics of the disease after brain injury can be estimated and analyzed.

6. Acknowledgments

Authors would like to express gratitude to Ilya Komoltsev and Ivan Kershner for providing the dataset. The research was supported by the Russian Scientific Foundation, project No. 16-11-10258.

7. References

[1] R. D'Ambrosio, E. Perucca Epilepsy after head injury / Current Opinion in Neurology, 2004 Dec; 17(6). — P. 731-735.

[2] M. Dichter Posttraumatic epilepsy: the challenge of translating discoveries in the laboratory to pathways to a cure / Epilepsia, 2009; 50 Supp. 2. — 41 p.

[3] I.A. Kershner, Yu.V. Obukhov, I.G. Komoltsev Method of automated epileptiform seizures and sleep spindles delection in the wavelet spectrogram of rats in EEG / 3rd International conference "Information Technology and Nanotechnology 2017". — Samara, 2017. - P. 104-109.