Prediction and detection of epileptic seizures

Gonalo Duarte, Tadeu Ferreira

November 13, 2017

Abstract

The aim of this work is to create a classifier using Neural Networks in order to predict and detect epileptic seizures. Our object of study was an EEG from two different patients presenting 29 features. Knowing that a seizure is composed by 3 parts, we developed two different neural networks, one for detecting and one for predicting testing different architectures and varying different parameters.

1 Introduction

EEG involves recording and analysis of electrical signals generated by the brain. It is an important clinical tool for diagnosing and monitoring of neurological disorders related to epilepsy. Epilepsy is characterized by sudden recurrent and transient disturbances of mental functions and/or movement of the body that results from excessive discharging of groups of brain cells. This work describes the prediction and detection of this disturbances using Neural Networks.

2 Methods

2.1 Data Set

The data set was provided by the professor and was extracted by Doctor Mojatba Bandarabadi from two patients: 54802 and 109602. It consists of an EEG with 29 features obtained in each second. In addition a Target vector was also provided, which identifies when the seizures happen (represented with 1's). Finally, it was necessary a procedure two identified all four phases of an epileptic seizure that will be explained in detail in the next section.

2.2 Pre-Processing and Data Balance

First it was necessary identify correctly the phase inter-ictal (when there is no seizure), pre-ictal (considered 600 seconds before the seizure), ictal (seizure) and finally pos-ictal (considered 300 seconds after the seizure ended). Once identified all phases, modifications were made in the Target vector. Since the Neural Network only accepts 0s or 1s, inter-ictal became [1;0;0;0], pre-ictal [0;1;0;0], ictal [0;0;1;0] and pos-ictal [0;0;0;1].

Due to the non-balancing of the existing data quantity of the respective classes, belonging to about 90% of the samples to the inter-ictal period, the division of the data set was not a simple proportional division. Down-sampling of the inter-ictal phase was necessary in order that the number of inter-ictal was about the same as the sum of pre-ictal, ictal and pos-ictal. In order to be able to generalize as much as possible, with a greater representation, the data were divided into 70% for Training and Validation and 30% for testing.

2.3 Neural Networks

2.3.1 Strategy

The proposed strategy was specializing the neural network for the correct detection and prediction of the seizures separately. It was tested two methods:

- the first method consisted in changing the Ictal Vector (detecting) from [0;0;1;0] to [1;0] and the other phases to [0;1]. In the same way, the Pre-Ictal vector (predicting) from [0;1;0;0] to [1;0] and the other phases to [0;1]. This way the classification would be only between two classes;

- the second method would consider all four classes but, depending of specialization of the neural network, apply bigger weights for Pre-Ictal in predicting and Ictal for detecting. In addition different architectures and parameters were tested using this two strategies, which will be explained in the next section.

2.3.2 Architectures

In this work it was used two different architectures FeedForward and Recurrent. FeedFoward is the simplest neural network, since the data only move in one direction; that is, they are introduced by the initial layer, treated by the layers hidden and exposed by the outer layer. In the Recurrent, the layers are linked in a loop, which makes the network dynamic. Thus, each layer is linked to the previous one with a delay, except the last, because it is the exposure layer. With these connections, the neural network has internal memory which allows a better prediction regarding a neural network with several simple layers.

In addition, Incremental and Batch Learning were tested in this work. In Incremental Learning the weights and bias are updated after each input is presented. The learning methods used were the gradient rule (matlab function learngd) and Hebb rule (matlab function learnh) and were iterative used in a cyclic order (matlab function trainc). In Batch Learning the weights and biases are updated at the end of an entire pass through the input data. Gradient Descent (matlab function traingd) and Scaled Conjugate Gradient (matlab function trainscg) were tested.

Furthermore, the number of hidden layers and neurons are tested. More hidden layers adds dimensionally allowing to solve more complex problems. Adding more neurons, it adds more parameters making it difficult to optimize the solution. More hidden layers and neurons are not always synonymous of better results.

Finally, the three transfer functions were used: Logarithmic Sigmoid (logsig), Linear (purelin) and Hard Limit (hardlim).

2.3.3 GPU

GPUs was mostly used to train all the networks (given the size of the data, the complexity of some networks and the number of epochs run we greatly recommend to run it on GPU if that option is present) , and see which parameters were revealing the best results.

2.4 Postprocessing

Knowing that the errors and the performance are computed point by point, we evaluated two hypothesis: detecting of the classes point by point; Or in windows of 10 points, count at least 5 points being classified as pre-ictal in predicting and ictal if is detecting.

2.5 GUI

In this section it will be briefly explained the GUI presented in this work. The GUI is divided in training a new network with the chosen patient and parameters and, testing a trained network in a chosen patient. All the parameters previously explained can be chosen in a very intuitive way to train the desired network. In short, GUI allows the user to train and test neural networks in a simple way. The GUI created in this work is presented in Figure 1.

3 Results

In order to evaluate the performance of the classifiers, the function *displayResults.m* presents in a table the sensitivity and prediction for both detection and predicting and also for the classifications point-by-point and 5 in 10 points. The best results are presented in the next tables.

4 Discussion

Developing Neural Networks to detect and predict epileptic seizures is real difficult task. The big variability of intra and inter patient makes the task difficult. The initial idea was to develop a the classifier with one patient and than test with the other, imitating real life procedures of developing a classifier Nevertheless, having only data of two patients, isn't enough to represent not a small part of all world representative. We decide to use the patient 109602 instead of 54802 because it seem to have a better total representation, although 54802 have more seizures.

Many parameters were tested. Observing table representing the results the architecture Recurrent is more present in the best results since they tend to have memory jut like the human brain and could be a better approximation of the in-vivo conditions. The structures [50] and [10 10] have the same number of best results but [10 10] seems to have higher results. This proves that more hidden layers are cable of solving more complex data. Logsig showed to be the activation function with more best results but Purelin have only one less and with the same performance of Logsig. Regarding Train functions is seem obvious that Trainscg is the best function. Trainc was also tested but, incremental learning showed to be deeply

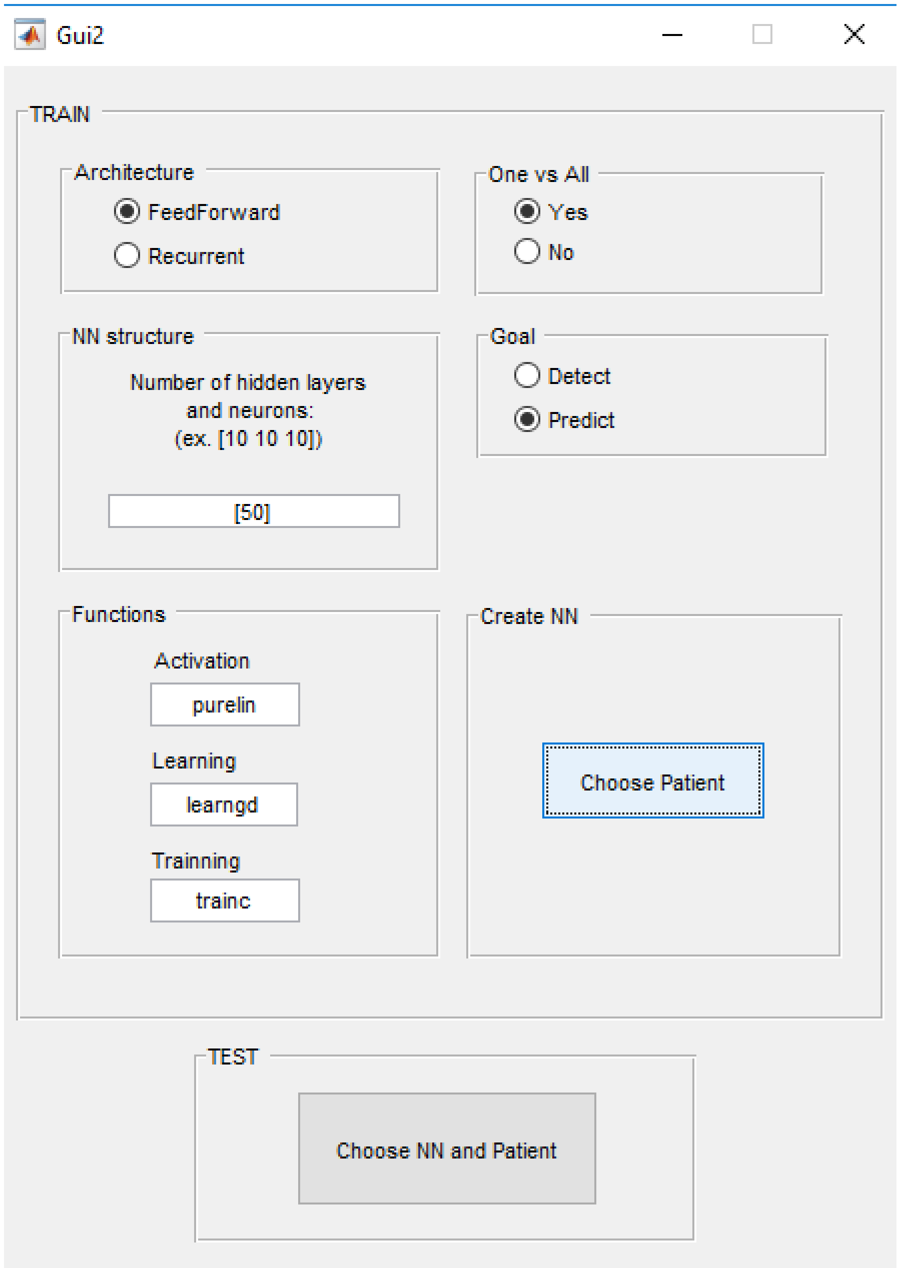

Figure 1: GUI - Example

heavy and with very low results. Scaled Conjugate Gradient Algorithm, a second order training algorithm for training of neural network, it provides faster training with excellent test efficiency. Finally the strategy of classification for only two classes seem to be best with only 2 best results for the All strategy. It seems that is easy for the Neural network to classify between two classes instead of four.

The extremist results of "100-0" or "0 -100" mean that classes are difficult to distinguish by parts of networks and that the variation of the parameters, in a general way makes the neural network consider everything positive or everything negative.

	Predict			
ID	**Sensitivity (%)**	**Specificity(%)**	**5/ 10 Sensitivity(%)**	**5/ 10 Specificity(%)**
1	0,08	99,42	96,77	1,38
4	0,26	98,87	96,77	1,38
6	0,01	99,85	70,97	40,49
17	0	99,95	29,03	87,41
20	0	99,97	41,94	67,71
49	0,09	99,51	96,77	1,38
50	0,18	99,21	96,77	1,38
62	1,51	99,1	77,42	23,37
66	0	99,98	19,35	93,1
73	0,12	99,77	41,62	41,77
54	0,01	99,89	70,97	40,5
57	0,46	99,7	61,31	50,53

Table 1: Parameters present in following order: architecture, structure, activation function, Train function, One vs All, Goal. **1** - FeedForward, [50], Logsig, Trainscg, Yes, Detect **4** - FeedForward, [10 100 10], Logsig, Yes, Detect **6**- FeedForward, [10 10], Purelin, Trainscg, Yes, Detect **17**- FeedForward, [50], Purelin, Trainscg, Yes, Predict **20** - FeedForward, [10 100, 10], Purelin, Trainscg, Yes, Predict **49** - Recurrent, [50], Logsig, Trainscg, Yes, Detect **50** - Recurrent [10 10], Logsig, Trainscg, Yes, Detect **54** - Recurrent, [10 10], Purelin, Trainscg, Yes, Detect **57** - Recurrent, [50], Hardlim, Trainscg, Yes, Detect **62** - Recurrent, [10 10], Logsig, Trainscg, Yes, Predict **66** - Recurrent, [10 10], Purelin, Trainscg, Yes, Predict **73** - Recurrent, [50], Logsig, Trainscg, No, Detect

In this work, two specialized classifier were developed: one for detecting and other for predicting. When the goal is predicting the classifier shows to have a very low sensitivity and a high specificity point by point. For the approach of finding 5 in 10 points shows a much higher sensitivity but a lower specificity. The pre-ictal class has more variability and doing block analysis gives better results. When the goal is detect, specificity point by point is very high and sensitivity is much higher for classifier specialized indirecting than predicting. This proves that the classifier is capable of distinguish very well ictal classes from others. The block approach seems to not work very well with detecting goal representing low sensitivity and mainly specificity unlike predicting for detecting goal. This proves that ictal classes have a very low variability.

5 Conclusion

In a general case it seems that the classification problem is directly dependent of the data where in different patient, one have better train performance that the other. In the future the train should be done with a higher number of patients. Data with only one patient isn't enough for world variability. Another solution for the future is a better extraction of features or another kind of analysis. It is possible that the solution is found in the pre-processing. Finally, the results were lower than expected, far to apply in clinical environment. The

	Detect			
ID	**Sensitivity (%)**	**Specificity(%)**	**5/ 10 Sensitivity(%)**	**5/ 10 Specificity(%)**
1	29,2	99,65	98,84	1,45
4	36,16	99,16	99	1,45
6	8,57	99,92	64,11	40,11
17	1,81	99,96	25,91	86,9
20	1,43	99,98	33,78	67,36
49	19,65	99,67	98,73	1,45
50	27,5	99,43	98,76	1,45
62	5,04	99,11	74,49	23,32
66	0,81	99,99	18,95	92,73
73	4,68	97,77	85,21	1,33
54	7,12	99,94	63,93	40,11
57	5,9	99,73	54,46	50,13

Table 2: Parameters present in following order: architecture, structure, activation function, Train function, One vs All, Goal. **1** - FeedForward, [50], Logsig, Trainscg, Yes, Detect **4** - FeedForward, [10 100 10], Logsig, Yes, Detect **6**- FeedForward, [10 10], Purelin, Trainscg, Yes, Detect **17**- FeedForward, [50], Purelin, Trainscg, Yes, Predict **20** - FeedForward, [10 100, 10], Purelin, Trainscg, Yes, Predict **49** - Recurrent, [50], Logsig, Trainscg, Yes, Detect **50** - Recurrent [10 10], Logsig, Trainscg, Yes, Detect **54** - Recurrent, [10 10], Purelin, Trainscg, Yes, Detect **57** - Recurrent, [50], Hardlim, Trainscg, Yes, Detect **62** - Recurrent, [10 10], Logsig, Trainscg, Yes, Predict **66** - Recurrent, [10 10], Purelin, Trainscg, Yes, Predict **73** - Recurrent, [50], Logsig, Trainscg, No, Detect

final conditions should be development of an algorithm that could specialize for a specific patients, preventing in a safe time when the seizure is going to happen.