Tarnu Ilie-Cristian 237

Report 3

Overview

The purpose of this paper is to compare the performance of different neural network architectures on the problem of epileptic seizure recognition. This is treated as a multi-label classification problem with 5 classes as provided by the dataset. 3 architectures were tested for this task, each being trained on the raw input signal and on the custom features that were extracted from the signal using discrete wavelet transform.

Dataset: University of Bonn EEG Dataset

This dataset contains EEG's from 500 subjects. The recordings were done for 23.5 seconds and have resulted in 4097 datapoints. This data is divided into 23 chunks which results in 178 data points in one second. In total there are 11500 rows, each with 178 points.

Each of these rows has been labeled according to the state of the subject (column 'y'), where the labels have the following meaning:

5 - Eyes open, means when they were recording the EEG signal of the brain the patient had their eyes open

4 - Eyes closed, means when they were recording the EEG signal the patient had their eyes closed

3 - The region of the tumor was identified and recording of the EEG activity was taken from the healthy brain area

2 - The EEG recording was taken from the area where the tumor was located

1 - Recording of seizure activity

Metrics

The metrics used for evaluating the performance of each model are:

$$Accuracy = \frac{(TP\ +\ TN)}{(TP\ +\ TN\ +\ FP\ +\ FN)}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TP + FN}$$

Together with the average loss for the test data, whose formula depends on the chosen loss function.

10-fold cross validation

Out of the 3 models, 2 were tested using 10-fold cross validation. By using this method we can get a more accurate representation of the models performance on another dataset. In this method the dataset is divided into 10 parts (more generally k), and then 1 part is chosen as the test data. The model is trained on the remaining 9 parts and the performance of the model for that particular “fold” is compared against the test data. This process is repeated until all parts have been chosen as test data. The final metrics are the averages of the metrics from each fold. For each fold the model is trained for 50 epochs.

Feature extraction

For feature extraction the discrete wavelet transform (DWT) method was used using the wavelets Daubechies 6, Symlets 6 and Coiflets2 as their output has the same length when applied to the same signal. When applying these wavelets (1 level deep) the signal is decomposed into 2 signals of equal length one being the approximate coefficients which result from applying the low pass filter from the wavelet and the other one being the detailed coefficients which result from applying the high pass filter from the wavelet. Whenever another wavelet was used to filter the data it will be mentioned.

General Designs

These design features appear in each model. The chosen loss function for this classification task was categorical cross entropy as the output from each model is a vector of length 5, denoting the probability of each class. As for the optimizer Adam with default parameters was chosen, because it appeared to decrease the loss faster than other optimizers. For the final classification each model uses 2-3 fully connected layers, where the last layers uses the softmax activation function and the others use sigmoid.

Model 1 – FCNN

The first model consists of a sequence of fully connected (also called dense) layers. There are 3 dense layers with the following number of neurons: 300, 100, 20, each layer using sigmoid as activation function. After these layers there is a dropout layer with 50% chance. This is used so that model does not overfit the test data. The final layer is the output layer and it is a dense layer with 5 neurons with sofmax as its activation function.

In the case where raw signal is passed to the model, the input consists of a 1 dimensional vector of length 178. Using this as input the model has a 50% average accuracy and an average loss of 1.193. When filtering the input only the detailed coefficients from the Daubechies 6 wavelet transformation were passed to the model. This however decreased the model's average accuracy by 2%, so the accuracy in this case is 48% and the average loss is 3.017.

Model 2 – CNN

This model is composed of 3 1D convolutional layers, each having different filter number and different kernel sizes. All convolutional layers use ReLU (rectified linear unit) activation function and use a stride of 1. The first convolution layer has 32 filters with a kernel size of 15, the second has 16 filters with kernel size 10 and the last convolutional layer has 8 filters with kernel size of 5. After these convolution layers a 1D max pooling layer is applied with a pool size of 2 in order to reduce the dimensions of the output from the convolutional layers. The reduced output is then flattened in order to be passed to the classification part of the model, which consists of 2 dense layers, separated by a dropout layer with 50% bias. The first dense layer has 40 neurons and uses sigmoid activation while the other one being the output layer has 5 neurons with softmax activtation.

The input layer for this model when using raw signals is also 1D vector of length 178, that means that the input for the first convolutional layer is 1 channel with 178 datapoints. The average accuracy in this case is 74%, having a sensitivity of 95% for class 1 which denotes seizure activity. When filtering the input into 6 signals using DWT with each of the above mentioned wavelets the accuracy drops to about 48% which and the sensitivity for each class is much lower. The average loss in this case is 1.132, while the raw signal input gives a loss of only 0.549.

Model 3 – RNN

This model has the least amount of layers compared with the other ones. This model, besides input and classification layer has only a GRU layer with 64 units. The GRU layer uses tanh as the activation function and sigmoid for the recurrent activation. The classification is the same as in all other cases and uses the same activations.

As this model takes a long time to train, only one fold was used to test the model. The model was trained for 50 epochs and had the following result: When not filtering the signal the model achieved accuracy of 71% and a sensitivity of over 90% in classes 1 (seizure activity), 4 (eyes closed - normal) and 5(eyes open - normal), which is higher than any other model. The same decrease in accuracy appeared here as well when first filtering out the input data. The filtered data consists of 6 vectors of the same size, which represent the output from the DWT of the wavelets: Daubechies 6, Symlets 6 and Coiflets2. The accuracy in this case was 66%.