Epileptic Seizure Detection from EEG Signals with Autoencoded Features

Tuan Nguyen

December 19, 2017

1 Domain Background

The domain of interest of this project belongs to physiological data such as electroencephalography (EEG), magnetoencephalography (MEG), electrocardiography (ECG) and the recorded signals from wearable devices. This project focuses on the EEG signals, which capture activities of brain neurons during a period of time. Different kinds of EEG data has been recorded from humans, for instance from those at rest, sleep [10], during some periods of specific cognitive activities, or from patients with diseases such as Alzheimer's, Parkinson's, depression and epileptic seisures ([1] and references thereof).

This project studies the classification problem on an EEG data set recorded from healthy volunteers and patients having epileptic seizures; solutions for this problem could be used to assist with detecting patients with the disease from their brain activity signals. Previous work on this data set focused mainly on extracting hand-crafted features to be used for classification. As examples, Nigam and Graupe [11] extracted the spike amplitudes and frequency of the signals, feeding them into a neural network for classification; Guler et al. [5] applied wavelet transformation to the signals to extract features, and classification was performed using a neuro-fuzzy system; Kannathal et al. [8] extracted entropy-based features for classification. This project, on the other hand, explores a type of autoencoders (e.g., [2], [13], [3], [9]) in order to learn features representing the EEG signals and then use them for classification.

2 Problem Statement

This project aims to classify 1-second long EEG segments into one of the three classes: (1) healthy volunteers, (2) patients with epileptic seizures disease during seizure-free periods, and (3) patients with the disease during active seizure periods. It can be formally defined as follows.

- **Input:** training set $X = \langle \vec{x}_1, \vec{x}_2, ..., \vec{x}_N \rangle$ of N samples, where $\vec{x}_i$ is a vector of M electrical voltage values recorded in one second by an electrode at a specific point and time; target vector $y = \langle y_1, y_2, ..., y_N \rangle$ where $y_i \in \{1, 2, 3\}$ is the type of volunteers which the sample $\vec{x}_i$ belong to. In the data set [1], the targets $1, 2, 3$ respectively correspond to sets of volunteers $A + B$, $C + D$ and E.
- **Output:** a hypothesis f classifying a sample of M electrical voltage values into one of the volunteer classes $\{1, 2, 3\}$.

The formation of the training set X and target y is discussed in Section 3. This project explores one of the following autoencoders for learning features of the training set X: ordinary autoencoders [2], denoising autoencoder [13], sparse autoencoder[3] and variational autoencoders [9]. The extracted features are then tested with popular learning models (Decision Trees, K-Kearest Neighbors, Support Vector Machines and Feedforward Multi-layer Neural Networks) for classification (see Section 4). Several metrics are used to evaluate the classifiers, discussed in Section 6.

3 Data Sets and Inputs

The data set used in this study is provided with the work by Andrzejak et al. [1], recording brain electrical activity of five sets of human volunteers:

- Set A: healthy volunteers in relaxed state with eyes open.
- Set B: healthy volunteers in relaxed state with eyes closed.
- Set D: epilepsy patients during seizure-free periods; the electrodes were implanted to record brain activity from within the epileptogenic zone.

- Set C: epilepsy patients during seizure-free periods; the electrodes were implanted from the opposite hemisphere of the brain (compared to those in Set D).

- Set E: same patients with sets C and D but activity were recorded from all electrodes during active seizures.

For each set of volunteers above, the data set contains 100 single-channel EEG segments of 23.6-sec duration. For this project each of these segments is divided into smaller segments of one second durations, which are considered stationary for EEG data [11]. These one second durations together define the training set X in the problem statement, and the three sets of volunteers $A + B$, $C + D$ and E define the target vector y in the problem statement above.

4 Solution Statements

This project aims to use a type of autoencoders ([2], [13], [3], [9]) that can learn to extract features of the training set X for the purpose of classifying new sample of EEG segments. I propose to first try ordinary stacked autoencoders [2]; the other autoencoders will be explored depending on the performance of the first one on this data set.

An ordinary autoencoder [2] is a regular neural network which consists of the input layer, one hidden layer and the output layer such that the input and output layers have the same number of neurons. The first (lowest) autoencoder on the stack is trained to minimize the difference, defined for instance as the mean squared error, between the input batches of EEG segments and their output activations. The output activation of the hidden layer becomes the input to the second autoencoder on the stack, which is trained also to minimize the difference between its input and output layers. This process of adding more autoencoders continues in that fashion. Finally, the output layer of the entire stack is the output of the first autoencoder. The stacked autoencoders is therefore symmetrical with regard to the middle hidden layer: the first lower half acts as the encoder while the second higher half of the stack as the decoder. The output activations of the middle layer thus act as encoded features of the training set of EEG segments.

Depending on the performance of the above autoencoder in classifying the target classes of volunteers, I might also plan to train a stack of denoising

autoencoders [13] to extract features for the data set. At high level, this stack is much like the ordinary one except for the Gaussian noises added into the input layer, forcing the decoder to learn more meaningful features, potentially improve the performance in classification.

After the feature learning phase above, the output of the coding layer the network is considered the features representing the training set X. They are then used as input features of popular classification algorithms. This work considers the following classifiers: Decision Trees, K-Kearest Neighbors, Support Vector Machines and Feedforward Multi-layer Neural Networks. The hypothesis here is that the features learned automatically using autoencoders, together with one of these classifiers, provides comparable performance compared to the previous work (see Section 5).

5 Benchmark Model

Several previous work has been developed for the classification problem on the same data set [1]; the main contribution of most of these work, however, were on designing various hand-crafted features for the EEG signals. They might also be different from each other in the target classes of interest. As examples, Nigam and Graupe [11] used two features based on the spike information of the signals for further learning and detecting patients in set E (thus, binary classification problem). Kabir et al. [7] extracted statistical features based on "optimum allocation technique" to be used with logistic model trees for classifying volunteers into the three classes that are of interest of this project. Guler et al. [5] employed Lyapunov exponents based features for classification using recurrent neural networks where the target classes are three sets A, D and E. On another data set, Wulsin et al. [14] learned features of second-long segments of EEG signals using Deep Belief Networks [6] and classified them into "clinically significant" EEG classes. To the best of my knowledge, there was no previous work reporting the results of using autoencoders for the data set that this project is interested in.

The proposed method by Kabir et al. [7] on the data set of interest in this project outperforms many the state-of-the-art approaches using the same data set; among many other measures, the total accuracy of their approach was 95.3%. Since its target classes are the same with those of interest in this project, and their work represents a class of work using hand-crafted features, in my opinion it is reasonable to use the performance of their work as the

baseline for the approach to be explored in this project. The hypothesis is that by using autoencoders, it is possible to learn a set of features automatically for classifying EEG signals with comparable results. The next section discusses measures to evaluate performance of classifiers in this domain.

6 Evaluation Metrics

The quality of the returned hypothesis is evaluated with an independent test set, created by 20% of the given data set. The following measures are used to evaluate the performance of different classifiers:

- F1 score: this measure combines both the precision and recall as follows

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall}.$$

- Total accuracy: the total number of samples classified into the correct classes, which is still the most common measure used in practice.
- Classification time: as suggested by Wulsin et al. [14], this measure is important for applications monitoring EEG signals of patients in real time.

7 Project Design

The project will go through the following stages. (Although the purpose and the direction of the project are fixed, the steps during the actual implementation might change.)

7.1 Data Preparation

The data set [1] contains 100 channels (or vector of voltage values) for each of the five sets of volunteers. Each channel is a vector of 4097 values that are EEG signals recorded in 23.6 seconds. The data set will be preprocessed as follows.

- Each channel is divided into 23 segments with equal length, each contains 178 values. These segments are marked with the target value 0, 1 or 2 for the classes $A + B$, $C + D$ and E of the original channel.

- The resulting set of segments are splitted into training and test set (using StratifiedShuffleSplit function from Scikit-Learn).
- The resulting set of segments in the training set are then normalized so that the values of EEG signals are in between $[0, 1]$.

7.2 Designing and training individual autoencoder

There are several design choices to make in designing the autoencoders:

- Number of autoencoders in the stack: this is a hyperparameter that needs to be searched for (through for instance cross-validation). I plan to start with 2 to 4 autoencoders.
- Activation functions: sigmoid would be used for the output layer; for hidden layers, I plan to try the Exponential Linear Units (ELU) [4].
- Regularization: I plan to use L2 regularizer.

7.3 Classifying EEG segments using extracted features

Given the extracted features obtained from the previous stage, it is not clear which classifier would be the best in classifying EEG segments into the target classes. I propose to test the following classification algorithms: Decision Trees, K-Kearest Neighbors, Support Vector Machines and Feedforward Multi-layer Neural Networks (potentially with Dropout [12]).

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*, 64(6):061907, 2001.

[2] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pages 153–160, 2007.

[3] Y.-l. Boureau, Y. L. Cun, et al. Sparse feature learning for deep belief networks. In *Advances in neural information processing systems*, pages 1185–1192, 2008.

[4] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[5] N. F. Güler, E. D. Übeyli, and I. Güler. Recurrent neural networks employing lyapunov exponents for eeg signals classification. *Expert systems with applications*, 29(3):506–514, 2005.

[6] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[7] E. Kabir, Siuly, and Y. Zhang. Epileptic seizure detection from eeg signals using logistic model trees. *Brain informatics*, 3(2):93–100, 2016.

[8] N. Kannathal, M. L. Choo, U. R. Acharya, and P. Sadasivan. Entropies for detection of epilepsy in eeg. *Computer methods and programs in biomedicine*, 80(3):187–194, 2005.

[9] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[10] M. Längkvist, L. Karlsson, and A. Loutfi. Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems*, 2012:5, 2012.

[11] V. P. Nigam and D. Graupe. A neural-network-based detection of epilepsy. *Neurological Research*, 26(1):55–60, 2004.

[12] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958, 2014.

[13] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

[14] D. Wulsin, J. Gupta, R. Mani, J. Blanco, and B. Litt. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *Journal of neural engineering*, 8(3):036015, 2011.