EEG Four Class Motor Classification with Data-Augmented Convolutional Recurrent Neural Networks (CRNN) and Variational Autoencoder Convolutional Neural Networks (VAE-CNN)

Kellen Cheng
UID: 905-155-544
kellentan@gmail.com

Abstract

Electroencephalography (EEG) is a modality that can capture neuron signals from the brain's scalp area [3]. The experiment performs classifications on EEG data taken from the 2008 Brain-Computer Interaction (BCI) competition [1]. The two architectures under study were a data-augmented convolutional recurrent neural network (CRNN) as well as a data-augmented variational autoencoder convolutional neural network (VAE-CNN). Both architectures utilized the same dataset augmentation techniques, and were trained in three types of experiments: training on all patients and testing on all patients, training on all patients and testing on a single patient, training on one patient and testing on that patient. The study showcases the high performance of the data-augmented CRNN in classifying the four class EEG data, achieving an accuracy of over 70% on all three types of experiments. While the implementation of the VAE demonstrated a potential for higher accuracy, these numbers were not realized in this study due to technological and time constraints with the computer.

1. Introduction

The study compares the testing performance of two deep learning neural network architectures on the 2008 BCI competition EEG data, the data-augmented CRNN and data-augmented VAE-CNN. Depending on the experiment type, the models are trained on either all patients or on the first patient, and their test accuracy metrics are compared. The data itself is separated into four different motor classes: left hand, right hand, both feet, and tongue [2]. Additionally, each model in each experiment type is saved and stored after training epoch hyperparameter testing, in order to ensure best performance. Additionally, we also used the data-augmented CRNN to identify the relationship between testing accuracy and signal time.

1.1. Data-Augmented Technique

Before the EEG signal data is processed by any of the models, the data first undergoes common augmentation techniques, which are intended to boost overall model classification performance. The data is first duplicated, and then Gaussian noise is added to the duplicated data. This transforms the data from shape (2115, 22, 1000) into shape (4230, 22, 1000). This technique essentially doubles the amount of effective training data that the model has available, allowing each architecture to fit better to the data. Only after this process is completed can the models begin the training process.

1.2. Data-Augmented CRNN

Please refer to Figure 3 for the architecture of the data-augmented CRNN. The model architecture was chosen empirically after extensive testing, focusing on selecting architectures that were both robust, simple, yet highly accurate. The model consists of three two-dimensional convolution layers, each equipped with the exponential linear unit (ELU) activation function, a two-dimensional max pooling, batch normalization, and a dropout layer. Afterwards, there is a dense layer, followed by a uniquely doubly-stacked long short-term memory (LSTM) layer, and finally a fully-connected output layer. Unlike a regular convolutional neural network (CNN), the LSTM layer is an addition that helps the CNN base architecture better handle time series information, which happens to be the point of focus for the EEG data. The LSTM allows the CNN to combat vanishing/enormous gradients and store short-term information for an extended period of time. Uniquely to the data-augmented CRNN, the doubly-stacked LSTM layer is composed of two LSTM layers, to allow for more robust retention of short-term information.

1.3. Data-Augmented VAE-CNN

Please refer to Figure 4 for the architecture of the data-augmented VAE-CNN. The model architecture was chosen empirically, but was influenced by enormous time and technological constraints. As the experiments could not be run on a GPU device, this condition dictated that the VAE-CNN contained a very basic structure. By minimizing the amount of trainable parameters and VAE complexity, the experiment time duration became much more manageable, at the tradeoff of better accuracy.

In the architecture, the augmented dataset is passed into the VAE. On the encoder side, the data undergoes two two-dimensional convolution layers (also with the ELU activation function) and then a dense layer. Additionally, the z-vector is then sampled according to the calculated z-mean and z-variance. On the decoder side, the architecture reshapes the values before performing two convolutional transposes to generate data with the same dimensions as the inputs. The decoder output is then stored along with the input data as a form of additional data to our dataset.

The CNN architecture attached to the VAE is highly simple, primarily due to computing power restraints. It consists of two two-dimensional convolution layers, each equipped with the ELU activation function, a two-dimensional max pooling, batch normalization, and a dropout layer. Finally there is a fully connected layer. The VAE that is attached to the CNN allowed it to perform better than a traditional CNN on the EEG data, although the full potential of this accuracy could not be realized in the light of poor computing resources.

2. Results

2.1. All Patient Experiments

For testing on all patients, the EEG data is split into a training and validation dataset, which each model architecture used to train on. After the training phase, each model is then evaluated based off of their classification accuracy on all test patients. Please refer to Table 1 for comprehensive accuracy metrics.

2.2. Single Patient Experiments

For testing on a single patient (i.e. patient one for this study), the EEG data is first filtered to exclude training and validation data not originating from patient one. Additionally, the test data is also filtered to exclude data not generated by patient one. Only after this selection is complete can the dataset then be split into training and validation datasets, which each model uses to train on. After the training phase, each model is then evaluated based off of their classification accuracy on a single patient (specified to be patient one). Please refer to Table 2 for comprehensive accuracy metrics.

2.3. Single Patient Experiments (Trained on All Patients

For testing on a single patient (i.e. patient one for this study), the EEG data is first filtered to exclude test data not generated by patient one. However, unlike the single patient experiments detailed earlier, each model will make use of the full training and validation dataset (i.e. the models are trained on all patients). After the training phase, each model is then evaluated based off of their classification accuracy on patient one. Please refer to Table 3 for comprehensive accuracy metrics.

3. Discussion

3.1. Model Comparison

Between both architectures, they share the common Adam optimizer and measure the categorical cross-entropy loss. Additionally, the data augmentation techniques for preprocessing are also shared between the models. Due to the technological constraints, the CRNN architecture demonstrates a clear performance gain over the simple VAE-CNN model. From empirical observation of earlier training epochs, it can be extrapolated that the VAE-CNN could have had high classification accuracy, although not necessarily higher than the CRNN model. This is believed to be due to the doubly-stacked LSTM element in the CRNN model, which allows for more robust handling of the time series EEG data, whereas the model attached to the VAE is a basic CNN. What was noteworthy however, was the high classification accuracy from the VAE CNN on patient one, provided that the model was trained on all patients. It is possible that the VAE allowed for greater single subject classification, if it was exposed to all subjects. However, amongst other metrics, the superior model of the two is clearly the data-augmented CRNN.

3.2. Parameter Experimentation

From the base VAE and CRNN structures, both models were tuned empirically for best possible performance. Key metrics that were targeted included the number of time points used in each model, the number of training epochs, and the amount of extra training data to append. A specialized training method was developed in order to detect how many epochs were required for optimal testing accuracy (varied between both models). This helped ensure that the experiment identified the optimal point at which the model neither underfitted nor overfitted to the training dataset. Additionally, after visualization of the four different motor classes, it was revealed that the greatest distinctions were found in the first 500 time points, whereas the latter 500 time points were quite similar and revealed little separation between each class. Interestingly, from Figure 1, the empirical results demonstrate that the greatest accuracy is achieved utilizing the first 600 time points instead, although this may be skewed because of the small amount of training epochs implemented. As a result, for longer training duration, the datasets were also trimmed to crop out the latter 500 time points. Finally, the addition of extra training data was tuned to discover the amount that allowed for highest classification accuracy while working within the technological constraints. This amount was diminished to the point where the better model still outputted over 70% classification accuracy while decreasing runtime speeds.

Regarding the VAE-CNN architecture, the model was forced towards a basic structure due to technological constraints, mainly the lack of access to a dedicated GPU. From Figure 2, it is evident that despite the loss, it can be seen as steadily decreasing. The training process cuts off at 30 epochs simply because any longer would stall the machine to idle. However, from the trends illustrated, it is evident that there was high potential for the VAE-CNN to output high classification accuracy.

3.3. Training on All Patients v. Training on Patient One

For the data-augmented CRNN architecture, the empirical results illustrate slightly better accuracy metrics when the model was trained on all patients and tested on all patients. This makes sense as the training dataset encompasses all patients and thus allows for a comprehensive classification. For classification accuracy on patient one, whether the model was trained on patient one or all patients made very little difference. This might be because the CRNN is capable of generalizing the data purely from training on patient one, such that additional data from other patients made little difference in its classifications.

For the VAE CNN architecture, the empirical results clearly demonstrate better performance when utilizing the entirety of the training dataset, as opposed to data drawn from patient one only. This makes sense as the overall training dataset is more likely to encompass the features required for geeneralization. Interestingly, there is an incredible gain in classification accuracy on patient one when the model is trained on all patients. Potentially, the larger training dataset allows the model to be exposed to a wider range of data, which allows it much greater customizeability when it comes to classification accuracy on any one patient.

References

[1] Bci competition iv. https://www.bbci.de/competition/iv/goals.

[2] C. Brunner. Bci competition 2008 - graz data set a. 2008.

[3] P. B. J. Satheesh Kumar. Analysis of electroencephalography (eeg) signals and its categorization- a study. *Procedia Engineering*, 38(1):2525–2536, 2012.

Model	Classification Accuracy
Data-Augmented CRNN	**70.43%**
Data-Augmented VAE-CNN	49.44%

Table 1. Overall Testing Accuracy (N.B. Technological constraints hindered the VAE-CNN)

Model	Classification Accuracy
Data-Augmented CRNN	**70.00%**
Data-Augmented VAE-CNN	46.00%

Table 2. Patient One Testing Accuracy (N.B. Technological constraints hindered the VAE-CNN)

Model	Classification Accuracy
Data-Augmented CRNN	**70.00%**
Data-Augmented VAE-CNN	58.00%

Table 3. Patient One Testing Accuracy Trained on All Patients (N.B. Technological constraints hindered the VAE-CNN)

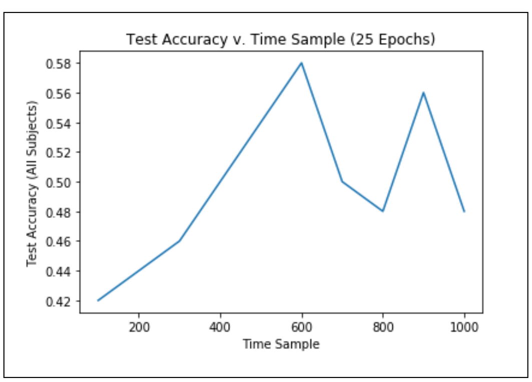

Figure 1. Test classification accuracy compared to the number of time points utilized in training.

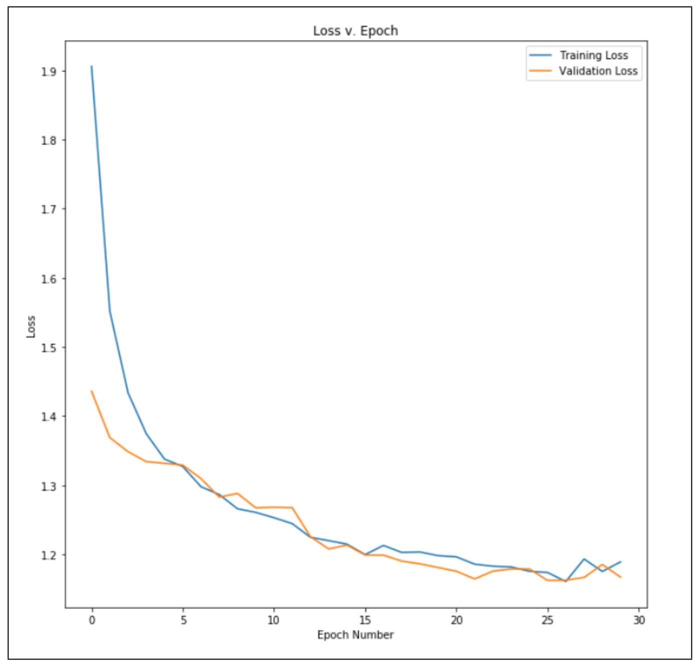

Figure 2. Loss of the VAE-CNN model over epochs.

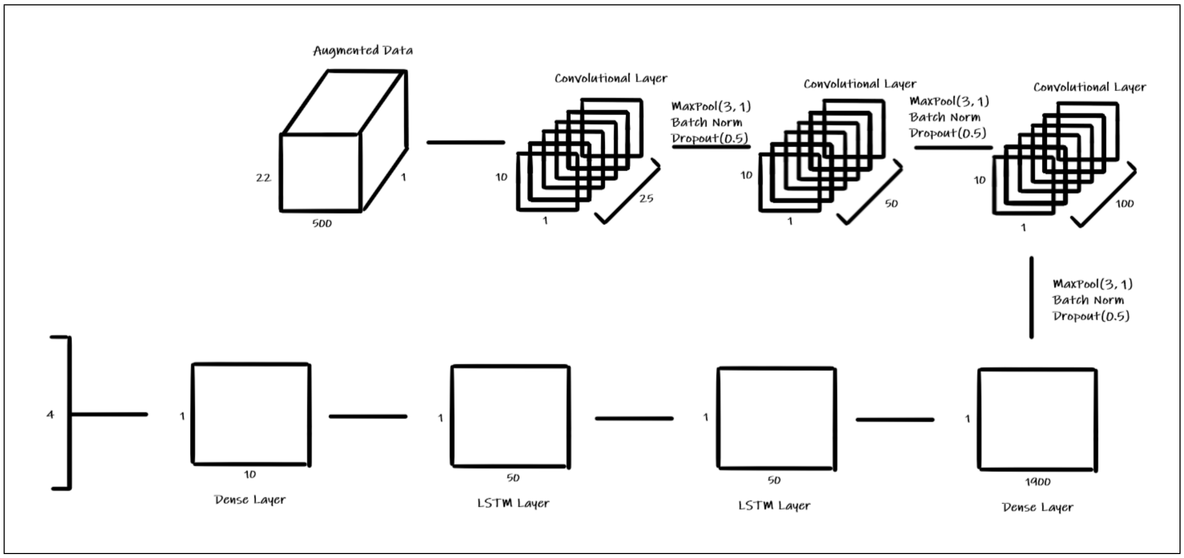

Figure 3. Architecture of CRNN

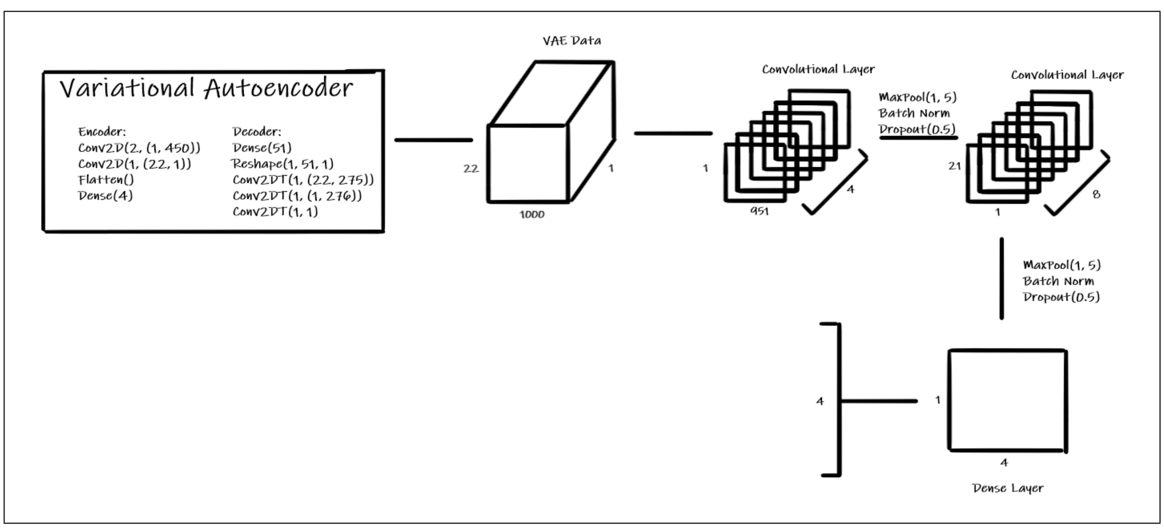

Figure 4. Architecture of VAE-CNN