Brain data domain adaption

Shayan Kousha[1,*], Demetres Kostas[1,2], and Frank Rudzicz[1,2,3]

[1]University of Toronto; Toronto, Canada
[2]Vector Institute; Toronto, Canada
[3]St. Michael's Hospital; Toronto, Canada
*shayan.kousha@mail.utoronto.ca

Abstract

Deep convolutional neural networks (CNN) have revolutionized machine learning techniques in computer vision and natural language processing. There is also an increasing interest among brain computer interface (BCI) researchers to use deep CNNs and adapt successful techniques from computer vision and NLP. However, many of these techniques are not immediately useful for electroencephalographic (EEG) recordings. One problem, in particular, is that inter-individual variability of brain signals makes the task of brain decoding challenging. In this paper, we present a method that uses gradient reversal, reconstruction loss, and a shared embedding space to partially remove the uniqueness from brain signals recorded from different individuals and improve the generalizability of CNN models. This method demonstrates that as long as signals are transferred to a common space properly, CNN models can be trained on multiple subjects and reach the accuracy close to the models trained on only one subject. To experiment with our method, we use ShallowConvNet classifier introduced by Schirrmeister et al. [1] and the BCI Competition IV Dataset 2a [4]. Our method increases the accuracy of ShallowConvNet model when it is trained on multiple subjects.

1. Introduction

Over the past few years, deep convolutional neural networks have affected almost all areas of machine learning such as computer vision and NLP. However, training accurate and generalizable neural net models for the purpose of decoding EEG brain signals has not been met with much success. Many of the CNN models designed for this purpose are only trained on data recorded from one subject [1]. Even though they perform well on the test set from the same subject, they do not generalize well to other subjects and the test accuracy would be much worse than the train accuracy. Additionally, attempts of training a model on data from multiple subjects have also not been successful as they fail to generalize when it comes to testing on subjects that are not a part of the training set and the results vary largely across different subjects [2].

The main reason that makes this task of designing a generalizable EEG decoder challenging is that brain signals are unique for every human. There is ongoing research investigating techniques to minimize (or possibly eliminate) the need for subject-specific information. However, many of these techniques do not apply CNNs in their methods and models [5, 6]. In this paper, we investigate techniques to reduce the uniqueness of brain signals in order to train robust systems. These techniques result in a model that outperforms a baseline model.

The problem of having data sampled from different domains with unique characteristics and making predictions solely based on features that cannot discriminate between the training domains is not specific to Motor Imagery EEG (MI-EEG). This problem is commonly referred to as domain adaptation. StarGAN [7] is an example of domain transformation in the field of computer vision. It adapted CycleGAN[8] to perform a multi-source domain adaptation task where it transfers a source domain to multiple target domains. For example, it transfers different hair colors or emotions for the same person. Commonly used methods for transferring source domains include adversarial domain adaptation [3, 9] and GAN-based domain adaptation [10, 11]. In this paper, we apply an adversarial domain adaptation technique called gradient reversal to improve the generalizability of the ShallowConvNet model.

We compare our result against four CNN models that are trained on the same database and their accuracies are reported as the average over all subjects (Fig. 2). The models include ShallowConvNet and DeepConvNet, which are introduced by Schirrmeister et al. [1], and EEGNet models introduced by Lawhern et. al. [2].

The remainder of this paper is organized as follows. After a brief review of the dataset in Sec. 2, we present in Sec. 3 the proposed method in detail. In Sec. 4, we present the performance of our method on the BCI IV 2A benchmark. Finally, we conclude the whole paper in Sec. 5.

2. Dataset

We use BCI competition IV 2A dataset which is publicly available. This dataset consists of EEG data from 9 subjects. The motor imagery tasks include movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Data is recorded in 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

We did not pre-process the signals much, except for bandpass-filtering the brain signals to 0-38 Hz and performing electrode-wise exponential moving standardization with a decay factor of 0.999 to compute exponential moving means and variances for each channel [1]. The moving means and variances were later used to standardize the continuous data.

Each row of the input matrix is an event that corresponds to one of our four classes. The data for each event includes the signals from 0.5 second before the start of the event to 4 seconds after the event.

We did not use any other data augmentation technique like the cropped training strategy [1] for brain signal augmentation mainly because our purpose was to train a model to perform an end-to-end EEG decoding with minimal preprocessing performed on the data. Additionally, we believe our model would not benefit from data augmentation. We will discuss it more in the discussion section.

3. Method

This section first introduces the architecture of our convolutional neural network, then presents the gradient reversal and compression strategy to remove noise and subject specific features, and finally details how the sub-models' parameters can be learned.

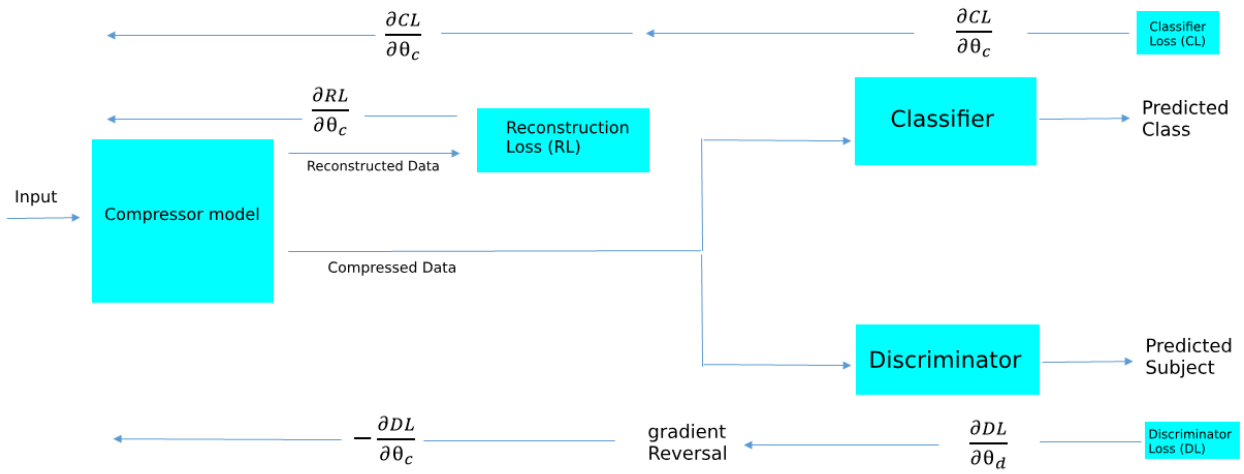

Figure 1: Model Architecture

3.1 Architecture

As shown in Fig. 1, the network consists of three parts, i.e., Compressor, Classifier and Discriminator models.

3.2 Classifier model

We used a brain decoder model named ShallowConvNet model introduced by Schirrmeister et al. [1] as our baseline and the classifier in our network. Given that the purpose of our research project is to find a common space between EEG brain data from different individuals in a way that it would be hard for classifiers to tell them apart, we found it unnecessary to design a new classifier model. We preferred to focus on the problem a hand which is the task of domain adaptation.

The input of this layer is the compressed data from the Compressor model. This model classifies each compressed event as one of the four motor imagery tasks.

3.3 Discriminator model

The Discriminator model is similar to ShallowConvNet. The purpose of this model is to tell the data of each individual apart. The gradient that updates the parameters of this using gradient reversal technique and passed to our Compressor model.

The discriminator consists of a layer for the purpose of gradient reversal, discussed in the next sub-section, as well as three convolutional layers. The first layer is a convolution across time and the second layer is a convolution across space (electrodes). Since there is no activation function in between the two convolutions, they could in principle be combined into one layer. However, using two layers implicitly regularizes the overall convolution by forcing a separation of the linear transformation into a combination of two (temporal and spatial) convolutions [1].

Similar to the Classifier model, the input comes from Compressor model. It predicts what subject the input event belongs to.

3.4 Gradient reversal

We refer to the gradient reversal layer mentioned in the previous section as a *flip_gradiant* layer [12] which is an implementation of gradient reversal [3] in TensorFlow. This layer has no effect on forward passes. However, it reverses the sign of gradients being passed from the Discriminator to the Compressor models. The reversed gradients confuses the Compressor layer and thus forces it to use features of brain data that are not unique to each subject.

3.5 Compressor model

The purpose of this model is to find a shared space between all the domains and compress the input signals in a way that their noise and unique features are removed without losing any of the features that are actually helpful for predicting the motor imagery tasks.
To accomplish this goal, we convolved the input data twice to compress the signals. Then we reconstructed it using two transpose convolutional layers. After processing the data, the model returns the compressed signals as well as the reconstructed ones. The compressed signals are passed to the Classifier and the Discriminator. The reconstructed signals are used to make sure our Compressor model compresses the data in a meaningful way.

	1	2	3	4	5	6	7	8	9	Average	SD
Test Accuracy (%)	57.98	44.09	64.9	51.38	53.81	52.43	60.76	58.68	63.54	56.38	6.22

Table 1: Test accuracy for all 9 subjects. To obtain the accuracies we train the model 9 times and used a different subject for testing purposes each time

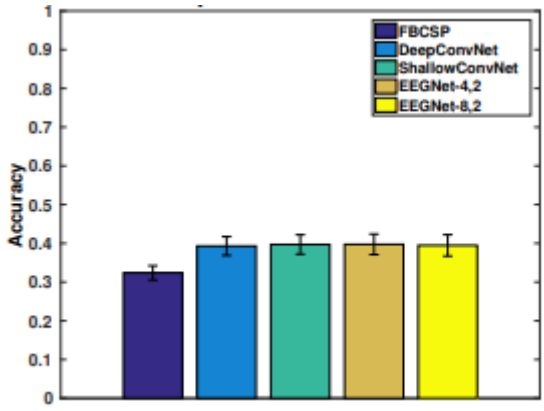

Figure 2: It is taken form Lawhern et. al. [2] representing a current attempt at cross-subject classification with the same dataset, averaged over all subjects. Error bars denote 2 standard errors of the mean.

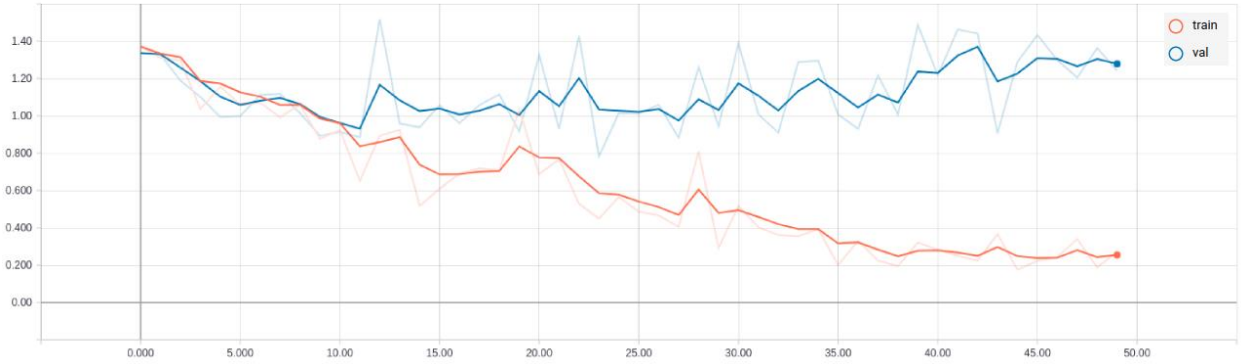

Figure 3: Classifier loss, trained on data of subject 1 to 8 for 50 epochs.

3.6 Learning

Learning a neural network model becomes complicated when techniques like gradient reversal are used. If the learning rate used for optimizing the Discriminator model is too high, then the error signals passed from the discriminator to the Compressor model might dominate the error signals it receives from the classifier. Therefore, to avoid this dominance of the discriminator, its learning rate must be set to a lower value than the learning rate of the classifier [3]. This is unfortunate because in almost all cases, the value assigned to the Discriminator model's learning rate is not the optimal one. Since the assigned learning rate is usually lower than the optimal one, the Compressor model might not get confused enough. Therefore, not all the features that makes each individual's brain signal unique will be removed completely from the data.

4. Experiment

Throughout the process of designing and modifying our network, we used subjects 1 to 8 for training and subject 9 for testing purposes. The EEG data from these 8 subjects is shuffled and split into training and validation set with the probability of 0.8. However, to demonstrate the effectiveness of our decoder mode, we trained the final version of the model, which its

hyperparameters are tuned, 9 times and used a different subject for testing purposes each time.

In order to record the data in table 1, we trained the model on the training set for 50 epochs and recorded checkpoints after each epoch. Then we used the checkpoints of the epoch with the smallest validation loss (Fig. 3) to classify the data in the test set.

Even though the recorded test accuracy for different subjects varies, all the accuracies are higher compared to other CNN brain data decoders when trained on multiple subjects from the BCI competition IV 2A dataset. Average test accuracy is 56.38% (SD=6.22%) which significantly outperforms the baseline models (Fig. 2) which their average test accuracies for cross-class classification are close to 40%.

5. Discussion

Even though the effectiveness of data augmentation techniques like the cropped training strategy in reducing overfitting has been proven multiple times [1, 13], we did not use such augmentation techniques because they did not increase the training and validation accuracy of the model. This was the case because our model partially removes the noise and subject specific features from the signals and reduces the need for data augmentation and hand designed features. Having too much data results in a slower training process with no gained benefit. Therefore, to avoid a slower training process and simplify our experiment, we decided to avoid data augmentation.

Our results show that the accuracy of deep convolutional neural networks trained on EEG brain signal can be improved significantly if the data are transferred to a shared space. Ultimately, we would like to develop a better classifier as we believe that there is room for improvement in the ShallowConvNet model. It is also worth noting that having brain signals from more subjects in the dataset should improve the generalizability. Therefore, we are eager to use a larger dataset than BCI VI 2a dataset.

Bibliography

[1] Schirrmeister, Robin Tibor, et al. “Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization.” *Human Brain Mapping*, vol. 38, no. 11, 2017, pp. 5391–5420.

[2] Lawhern, Vernon J, et al. “EEGNet: a Compact Convolutional Neural Network for EEG-Based Brain–Computer Interfaces.” *Journal of Neural Engineering*, vol. 15, no. 5, 2018, p. 056013.

[3] Ganin, Yaroslav, et al. “Domain-Adversarial Training of Neural Networks.” *Domain Adaptation in Computer Vision Applications Advances in Computer Vision and Pattern Recognition*, 2017, pp. 189–209.

[4] Tangermann, et al. “Review of the BCI Competition IV.” *Frontiers*, Frontiers, 30 Mar. 2012, www.frontiersin.org/articles/10.3389/fnins.2012.00055/full.

[5] Lotte, Fabien. “Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces.” *Proceedings of the IEEE*, vol. 103, no. 6, 2015, pp. 871–890.

[6] Waytowich, Nicholas R., et al. “Spectral Transfer Learning Using Information Geometry for a User-Independent Brain-Computer Interface.” *Frontiers in Neuroscience*, vol. 10, 2016.

[7] Choi, et al. “StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation” 24 Nov. 2017

[8] Zhu, Jun-Yan, et al. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.” *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017

[9] Wulfmeier, Markus, et al. “Addressing Appearance Change in Outdoor Robotics with Adversarial Domain Adaptation.” *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017

[10] Ghifary, Muhammad, et al. “Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation.” *Computer Vision – ECCV 2016 Lecture Notes in Computer Science*, 2016, pp. 597–613.

[11] Liu, et al. “Coupled Generative Adversarial Networks” 24 Jun. 2016

[12] Pumpikano. “Pumpikano/Tf-Dann.” *GitHub*, 28 Jan. 2018, github.com/pumpikano/tf-dann.

[13] Fedjaev. “Decoding EEG Brain Signals using Recurrent Neural Networks” 12 Dec. 2017