Learning from Heterogeneous EEG Signals with Differentiable Channel Reordering

Aaqib Saeed[1], David Grangier[2] Olivier Pietquin[2] & Neil Zeghidour[2]

[1]Eindhoven University of Technology, [2]Google Research

Overview

Convolutional networks have been widely adopted for EEG classification tasks (e.g., seizure detection) and require consistently ordered channels. However, datasets collected with different EEG headsets can vary in number and placement of channels. CHARM allows training a single model from such heterogeneous signals in an end-to-end manner, and transferring learned representations between headsets, opening the door for large-scale training over many EEG datasets.

Contributions:

- We introduce a novel framework CHAnnel Reordering Module (CHARM) for training a single model across varying EEG collections that can differ both in number and location of electrodes.
- Our differentiable module uses an attention mechanism on multichannel EEG signals, identifies the location of each channel from their content, and remaps them to a canonical order.
- We perform experiments on four EEG classification datasets for the tasks of seizure, artifact detection, and abnormal EEG recordings. We demonstrate the efficacy of via simulated shuffling and masking of input channels.
- We also propose a *channel masking and shuffling augmentation* strategy for multi-channel input to improve robustness of standard models.

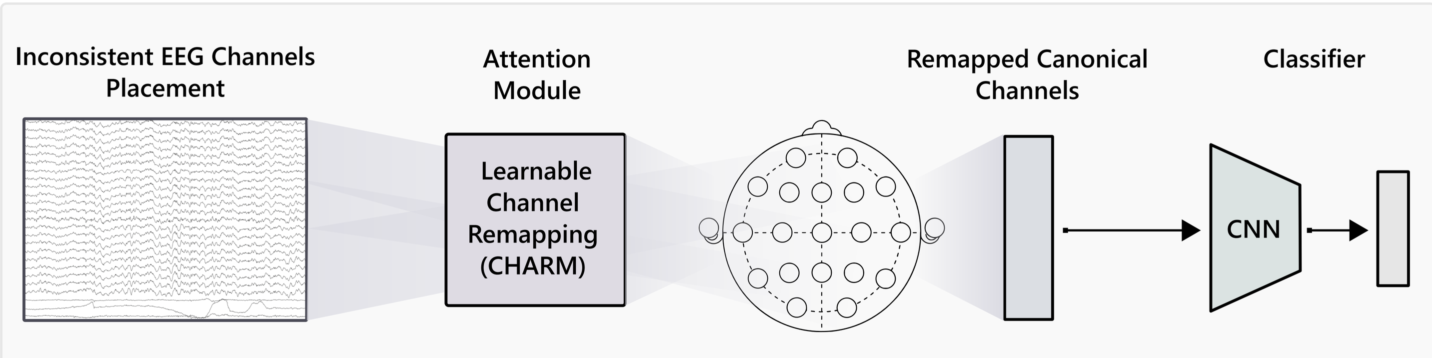
Learnable Channel Remapping

Our differentiable reordering module outputs a soft reordering matrix p. The input signal $\mathbf{x} \in \mathbb{R}^{N \times T}$ is a recording over N channels for a duration T. Considering M canonical channels, the reordering matrix $p(\mathbf{x})$ is an $N \times M$ matrix. Precisely, each canonical output is estimated as a weighted sum of the input channels, i.e.,

$$\hat{\mathbf{x}}_{i,t} = \sum_{j=1}^{N} p(\mathbf{x})_{i,j}\, \mathbf{x}_{j,t}\,, \quad i = 1, \ldots, M, \quad t = 1, \ldots, T.$$

$\hat{\mathbf{x}} \in \mathbb{R}^{M \times T}$ then serves as input to a standard neural network expecting a consistent input ordering across data samples. We consider three variants of our reordering method.

Overview of CHARM with a 1D convolutional classifier.



Convolutional Reordering

CHARM-base represents the signal of each channel as a vector,

$$h_i = m^{\text{conv}}(\mathbf{x}_{i,:})$$

where m^{conv} composes of a 1-D convolution layer with d filters and an aggregation operation to map a single-channel temporal signal into a fixed dimensional vector of dimension d.

Each vector is compared to learned embeddings of dimension d that represent each of the M canonical channels, $c \in \mathbb{R}^{M \times d}$, yielding the matrix p for channel remapping,

$$p_{i,j} = \text{softreorder}(c, h)_{i,j} = \frac{\exp(c_i \cdot h_j)}{\sum_{j'} \exp(c_i \cdot h_{j'})}.$$

Attentive Reordering

CHARM-CKV builds upon residual attention [1]. We build a *query vector* representing each input channel as $q_i = m^{\text{conv}}(\mathbf{x}_{i,:})$. Each query vector *attends* over canonical channels, i.e. each input channel query is mapped to a weighted sum of canonical channel *value vectors* according to their similarity to canonical *key vectors*,

$$h_i = \sum_j a_{i,j} v_j \quad \text{where} \quad a_{i,j} = \frac{\exp(q_i \cdot k_j)}{\sum_{j'} \exp(q_i \cdot k_{j'})},$$

and k, v are key and value embeddings representing the canonical channels. These layers can be stacked after a residual connection,

$$q_i^{l+1} = \text{layernorm}(q_i^l + \text{mlp}^l(h_i^l))$$

where q^l, h^l are the query and attentive representations of layer l. At the last layer, we compute $p = \text{softreorder}(c, q^l)$.

CHARM-CQ reverses the role of canonical and input channels, using input keys and values and relying on canonical queries. Input keys and values result from an independent channel-wise convolution with 1D filters,

$$(k_i, v_i) = m^{\text{conv}}(\mathbf{x}_{i,:})$$

and initial canonical queries q are learned embeddings. The reordering matrix becomes, $p = \text{softreorder}(q^l, k^l)$.

Generalizing to Shuffled and Masked Channels

CHARM improves robustness to missing and shuffled channels significantly even when more than 50% of the channels are missing.

Dataset	Method	Clean	Noisy	Noisy-25%	Noisy-50%	Noisy-75%
TUH Abnormal	Baseline	**0.830**	0.566	0.581	0.589	0.548
	CHARM-base	0.766	0.742	0.760	0.743	0.731
	CHARM-CKV	0.772	**0.751**	**0.767**	**0.756**	**0.747**
	CHARM-CQ	0.75	0.743	0.744	0.744	0.741
TUH Artifact	Baseline	**0.711**	0.243	0.245	0.176	0.263
	CHARM-base	0.618	0.514	**0.566**	0.517	0.466
	CHARM-CKV	0.628	0.481	0.521	0.491	0.452
	CHARM-CQ	0.607	**0.524**	0.538	**0.531**	**0.505**
TUH Seizure	Baseline	**0.950**	0.289	0.368	0.297	0.171
	CHARM-base	0.906	0.663	0.818	0.693	0.502
	CHARM-CKV	0.912	0.713	0.842	0.756	0.591
	CHARM-CQ	0.890	**0.770**	**0.857**	**0.808**	**0.704**
CHB-MIT	Baseline	**0.658**	0.371	0.296	0.363	0.439
	CHARM-base	0.554	0.504	0.523	0.503	0.487
	CHARM-CKV	0.562	0.518	0.543	0.529	0.504
	CHARM-CQ	0.576	**0.541**	**0.560**	**0.550**	**0.530**

Performance in Structured Masking Conditions

In a more tangible setting than random masking on a TUH Seizure dataset, we show CHARM performs well even when only a subset of channels from a specific half of the brain is active.

Augmentation	Method	Clean	Horizontal		Vertical	
			Group$_A$	**Group$_B$**	**Group$_A$**	**Group$_B$**
None	Baseline	**0.951**	0.631	0.430	0.486	0.586
	CHARM-CKV	0.900	0.790	0.600	0.683	0.762
	CHARM-CQ	0.899	**0.839**	**0.707**	**0.751**	**0.824**
CMSAugment	Baseline	**0.873**	0.823	0.762	0.779	0.783
	CHARM-CKV	0.829	**0.850**	**0.778**	**0.794**	**0.853**
	CHARM-CQ	0.734	0.750	0.702	0.711	0.739

Evaluation in Transfer Learning

We transfer learned representations from TUH Seizure to CHB MIT dataset which is collected from a different EEG headset.

Method	**Fixed**	**Fine-tuned**
Baseline (in-domain)	0.891 ± 0.038	
Baseline (transfer)	0.757 ± 0.004	$\mathbf{0.963 \pm 0.015}$
CHARM-CKV	$\mathbf{0.805 \pm 0.008}$	0.942 ± 0.020
CHARM-CQ	0.795 ± 0.006	0.915 ± 0.026

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need", *arXiv preprint arXiv:1706.03762*, 2017.