Inconsistent EEG Channels

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.

Overview

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

\[ h_i = m^{\text{raw}}(x_{c_i}) \]

where \( m^{\text{raw}} \) 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,

\[ p_{j,i} = \text{softreorder}(c, h), h_{j,i} = \exp(c_i \cdot h_j) \sum_j \exp(c_i \cdot h_j) \]

Learnable Channel Remapping

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

\[ x_{c_j} = \sum_{i=1}^{N} p(x)_{j,i} x_{c_i}, \quad i = 1, \ldots, M, \quad t = 1, \ldots, T. \]

\( x \in \mathbb{R}^{M \times T} \) 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 1D convolutional classifier.

Convolutional Reordering

CHARM-CKV builds upon residual attention [1]. We build a query vector representing each input channel as \( q_i = m^{\text{raw}}(x_{c_i}) \). 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,

\[ h_i = \sum_j a_{i,j}^{q_k} \text{ where } a_{i,j}^{q_k} = \frac{\exp(q_i \cdot k_j)}{\sum_j \exp(q_i \cdot k_j)} \]

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

\[ q^{l+1} = \text{layernorm}(q^{l} + \text{mlp}(h^{l})) \]

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

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,

\[ (k, v)_l = m^{\text{raw}}(x_{c_l}) \]

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

Attentive Reordering

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

Generalizing to Shuffled and Masked Channels

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.

Evaluation in Transfer Learning

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

References