

Cross-attention and CCA in EEG

Eduard Vladimirov, Daniil Kazachkov, Vadim Strijov

2 декабря 2024 г.

1 Abstract

На сегодняшний день работа с мультимодальными данными набирает всё большую популярность: учет взаимосвязей между ними улучшает качество предсказания. В этой статье мы предлагаем новую архитектуру, использующую преимущества алгоритма Canonical Correlation Analysis (CCA) и механизма Attention. Ниже будет показано, что CCA - частный случай Attention, а значит, мультимодальность можно встроить внутрь фреймворка Attention. Работа полученной модели иллюстрируется на задаче классификации удара теннисного мяча по датасету Real World Table Tennis.

keywords : CCA, Attention, BCI, online-classification.

2 Introduction

Мультимодальность - мощный инструмент для улучшения качества ответов модели [What Makes Multi-modal Learning Better than Single]. Канонический корреляционный анализ (CCA) [A Survey on Canonical Correlation Analysis] является очень популярным статистическим методом, снижения размерности двух множеств данных, при котором корреляция между парными переменными в общем подпространстве взаимно максимизируется. В таких работах как [Correlational Neural Networks], [A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets] авторы показали, что он улучшает качество в задачах сопоставления событий. Однако CCA может моделировать лишь линейные зависимости.

Существует несколько подходов по улучшению CCA, например, Kernel-CCA и Deep-CCA, каждый из которых имеет свои преимущества. **Тут**

вставка про такие улучшения, как Kernel CCA, Deep CCA... и их показатели.

В противоположность этому, механизм Attention находит сложные, нелинейные зависимости. Усовершенствование cross-attention позволит лучше отсеивать информацию, снизит размерность пространства и тем самым ускорит вычисления.

Ставя перед собой цель использовать преимущества каждого метода, мы представляем модель CCT: Canonical-Correlation Transformer. Архитектура у нее следующая: из пакета PyRiemann (мб заменим на CNN EEGNet) [ссылка] мы берем энкодер и преобразуем поданный на вход ЭЭГ сигнал в скрытое пространство. Далее в качестве механизма внимания используем ... (дополнить, когда станет понятно). Выход модели - вероятность события принадлежать одному из четырех классов: попадание по мячу (1), нейтральное событие (0), промах по мячу (-1), попадание, после которого успешных ударов не было (2).

Наша задача: как по данным ЭЭГ игрока в настольный теннис в режиме реального времени классифицировать момент удара - типичный пример из области Brain-Computer Interface (BCI), когда необходимо эффективно работать с данными разных модальностей и классифицировать события в онлайн режиме. В качестве датасета мы взяли предобработанные данные из "Real World Table Tennis"[1].

2.1 Related Works

- про нейрокомпьютерный интерфейс
- про онлайн-классификацию
- про emotional recognition с attnetion cca
- мб про аналогичные оффлайн-задачи

Нейрокомпьютерный интерфейс (Brain-Computer Interface, BCI) [2] считывает сигналы поверхности кортекса головного мозга, анализирует и переводит в команды исполняющей системы. Результатом измерений является временной ряд напряжений на электродах, который используется в задаче декодирования сигнала.

Одним из наиболее распространенных неинвазивных методов получения информации об электрической активности мозга является ЭЭГ.

Умение эффективно работать с подобными данными полезно во многих задачах: распознавание эмоций [ссылка], прогнозирование рецидивов болезни [ссылка], анализ сна [ссылка] и других. В этих примерах прослеживается и другая важная область: онлайн-классификация или stream-processing.

В этой работе авторы делали вот это.., использовали вот это. Явные методы, почему именно их, а не другие.

2.2 problem statement

3 CCT: Canonical-Correlation Transformer

3.1 Model

Мы продолжаем расширять область применимости CCT и представляем способ встраивания его в Attention.

- отдельно CCA
- отдельно Attention
- связь CCA + Attention (у Эдуарда из файла взять)

Canonical Correlation Analysis (CCA) - стандартный инструмент для выявления линейных зависимостей между двумя наборами данных [Canonical correlation analysis: An overview with application to learning methods]. Пусть нам дано множество векторов $X \in \mathbb{R}^{n_1 \times m}$ и $Y \in \mathbb{R}^{n_2 \times m}$, где m - количество векторов. Задача CCA - найти такие аффинные преобразования $\mathbf{W}_x, \mathbf{W}_y$, которые максимизируют корреляцию между X, Y в новом пространстве:

$$\begin{aligned}
\mathbf{W}_x^*, \mathbf{W}_y^* &= \arg \max_{\mathbf{W}_x, \mathbf{W}_y} \text{corr}(\mathbf{W}_x^\top X, \mathbf{W}_y^\top Y) \\
&= \arg \max_{\mathbf{W}_x, \mathbf{W}_y} \frac{\mathbf{W}_x^\top \hat{\mathbf{E}}[XY^\top] \mathbf{W}_y}{\sqrt{\mathbf{W}_x^\top \hat{\mathbf{E}}[XX^\top] \mathbf{W}_x \mathbf{W}_y^\top \hat{\mathbf{E}}[YY^\top] \mathbf{W}_y}} \\
&= \arg \max_{\mathbf{W}_x, \mathbf{W}_y} \frac{\mathbf{W}_x^\top C_{12} \mathbf{W}_y}{\sqrt{\mathbf{W}_x^\top C_{11} \mathbf{W}_x \mathbf{W}_y^\top C_{22} \mathbf{W}_y}},
\end{aligned} \tag{1}$$

где $\hat{\mathbf{E}}[f(\mathbf{x}, \mathbf{y})] = \frac{1}{m} \sum_{i=1}^m f(\mathbf{x}_i, \mathbf{y}_i)$, матрицы ковариации X и Y есть $C_{11} = \frac{1}{m} X X^\top \in \mathbb{R}^{n_1 \times n_1}$, $C_{22} = \frac{1}{m} Y Y^\top \in \mathbb{R}^{n_2 \times n_2}$, а матрица кросс-ковариации X, Y есть $C_{12} = \frac{1}{m} X Y^\top \in \mathbb{R}^{n_1 \times n_2}$.

Развивая идею [Learning Relationships between Text, Audio, and Video via Deep Canonical Correlation for Multimodal Language Analysis] для решения воспользуемся методом Singular Value Decomposition (SVD, Martin and Maes 1979) для $Z = C_{11}^{-1/2} C_{12} C_{22}^{-1/2}$ и получим матрицы U, S, V . Тогда

$$\begin{aligned} \mathbf{W}_x^* &= C_{11}^{-\frac{1}{2}} U = \left(\frac{1}{m} X X^\top \right)^{-\frac{1}{2}} U \\ \mathbf{W}_y^* &= C_{22}^{-\frac{1}{2}} V = \left(\frac{1}{m} Y Y^\top \right)^{-\frac{1}{2}} V \\ \text{corr}(\mathbf{W}_x^{\top*} X, \mathbf{W}_y^{\top*} Y) &= \text{trace}(Z^\top Z)^{\frac{1}{2}} \end{aligned} \quad (2)$$

В таких работах как [Learning Relationships between Text, Audio, and Video via Deep Canonical Correlation for Multimodal Language Analysis], [Deep Canonical Correlation Analysis] раскрыт подход использования глубоких сетей для обучения нелинейных преобразований двух наборов данных в пространство, в котором данные сильно скоррелированы. Мы же рассмотрим механизм внимания [Neural machine translation by jointly learning to align and translate], который используется для определения важности разных частей входных данных.

Механизм самовнимания определяется следующим образом:

$$\begin{aligned} \text{attn} : \mathbb{R}^{m \times d} \times \mathbb{R}^{m \times d} \times \mathbb{R}^{m \times d} &\longrightarrow \mathbb{R}^{m \times d} \\ \text{attn}(Q, K, V) &= \varphi \left(\frac{Q K^\top}{\sqrt{d}} \right) V \end{aligned} \quad (3)$$

where $Q, K, V \in \mathbb{R}^{m \times d}$ represent the queries, keys, and values, respectively, and $\varphi : \mathbb{R}^{m \times m} \longrightarrow \mathbb{R}^{m \times m}$ is row-wise applied nonlinear function, usually softmax. The dot product between Q and K determines the attention weights, which are normalized using the softmax function. The result is then applied to the values V to generate the output.

Self-attention applied to the input $X \in \mathbb{R}^{m \times n_1}$ is computed as:

$$\begin{aligned} \text{self-attn} : \mathbb{R}^{m \times n_1} &\longrightarrow \mathbb{R}^{m \times d} \\ \text{self-attn}(X) &= \text{attn}(X W_q, X W_k, X W_v) \end{aligned} \quad (4)$$

where $W_q, W_k, W_v \in \mathbb{R}^{n_1 \times d}$ — parameter matrices

In multihead attention, several attention heads are used in parallel, where each head computes its own attention weights and outputs. The outputs are then concatenated and linearly transformed by a weight matrix $W^Q \in \mathbb{R}^{p \cdot d \times d}$:

$$\text{multihead-attn}(Q, K, V) = [\text{head}_1, \dots, \text{head}_p]W^Q, \quad (5)$$

where $\text{head}_i = \text{self-attn}(X)$

Cross-attention, in contrast, involves attention between two different sets of inputs. It computes attention by using one set of inputs for queries $X_1 \in \mathbb{R}^{m \times d_1}$ and another set for keys and values $X_2 \in \mathbb{R}^{m \times d_2}$:

$$\text{cross-attn}(X_1, X_2) = \text{attn}(X_1W_q, X_2W_k, X_2W_v) \quad (6)$$

CCA and attention

Both CCA and attention mechanisms aim to find relationships between two sets of data. However, they differ significantly in their approach and applications:

Aspect	Attention	Canonical Correlation Analysis (CCA)
Goal	Identify relevant parts of input sequences	Receive embeddings in the same hidden space + dimensionality reduction
Similarity Measure	$A = \frac{1}{\sqrt{d}}QK^\top$ – attention matrix	$\text{tr}(A^\top S_{12}B)$, s.t. $A^\top S_{11}A = B^\top S_{22}B = I$
Optimization Goal	Minimize task-specific loss	$\max_{A,B} \text{corr}(A^\top X, B^\top Y)$

Таблица 1: Comparison of Attention Mechanisms and CCA

Note that $A^\top S_{12}B = \frac{1}{m}A^\top XY^\top B = \frac{1}{m}A^\top X (B^\top Y)^\top = \frac{1}{m}\widehat{Q}\widehat{K}^\top$. And it's quite similar to attention matrix formula $A = \frac{1}{\sqrt{d}}QK^\top$. Especially, in cross attention case, where Q is a linear transformation of X_1 and K is a linear transformation of X_2 :

Attn	Self-attn	Cross-attn	CCA	CCA-X	CCA-Y
Q	$W_Q^\top X$	$W_Q^\top X$	$A^\top X$	$S_{11}^{-\frac{1}{2}} X$	$S_{11}^{-\frac{1}{2}} X$
K	$W_K^\top X$	$W_K^\top Y$	$B^\top Y$	$S_{22}^{-\frac{1}{2}} Y$	$S_{22}^{-\frac{1}{2}} Y$
V	$W_V^\top X$	$W_V^\top Y$	I	$S_{11}^{-\frac{1}{2}} X$	$S_{22}^{-\frac{1}{2}} Y$
φ	softmax	softmax	Id	SVD $_U$	SVD $_V$

Таблица 2: United notation of CCA and attention

Let’s view in detail the CCA projection of X to latent space:

$$\begin{aligned}
\text{CCA}_{XY}(X) &= U^\top S_{11}^{-\frac{1}{2}} X = U^\top X_1 \\
\text{CCA}_{XY}(Y) &= V^\top S_{22}^{-\frac{1}{2}} Y = V^\top Y_1 \\
Z &= S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}} = \frac{1}{m} X_1 Y_1^\top
\end{aligned} \tag{7}$$

4 Experiments

[какие-то общие слова, если нужны] Например, про то, что пробо-
вали PyRiemann и трансформер из braincode.

4.1 Dataset Details

- что за датасет
- какая предобработка данных проводилась
- в каком виде данные подавались в модель

4.2 Training Details

Если возникнут какие-то проблемы или эвристики при обучении,
то пишем сюда

- На каких параметрах обучали сетку,

4.3 Experimental Results

Получилась вот такая точность и почему.

5 Appendix

5.1 CCA-Attention table

Приведем пример вывода значений таблицы для одного примера:

Список литературы

- [1] Amanda Studnicki and Daniel P. Ferris. Dual-layer electroencephalography data during real-world table tennis. *Data in Brief*, 52:110024, 2024.
- [2] Jerry J Shih, Dean J Krusienski, and Jonathan R Wolpaw. Brain-computer interfaces in medicine. In *Mayo Clinic Proceedings*, volume 87, pages 268–279. Elsevier, 2012.
- [3] Nick Martin and Hermine Maes. Multivariate analysis. *London, UK: Academic*, 1979.
- [4] Zhongkai Sun, Prathusha Sarma, William Sethares, and Yingyu Liang. Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8992–8999, 2020.
- [5] Yu-Ting Lan, Wei Liu, and Bao-Liang Lu. Multimodal emotion recognition using deep generalized canonical correlation analysis with an attention mechanism. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2020.
- [6] Yu Zhang Ehsan Adeli Qingyu Zhao Kilian M. Pohl Yixin Wang, Wei Peng. Brain-cognition fingerprinting via graph-gcca with contrastive learning. 2024.
- [7] Engin Erzin Ibrahim Shoer, Berkay Kopru. Role of audio in audio-visual video summarization. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. IEEE, 2023.
- [8] Bai Chenyu Pan Jiahui. Eeg-based emotion recognition via convolutional transformer with class confusion-aware attention. In

Proceedings of the Annual Meeting of the Cognitive Science Society,
volume 46. IEEE, 2024.