Cross-attention and CCA in EEG

Eduard Vladimirov, Daniil Kazachkov, Vadim Strijov

8 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

Современные задачи, связанные с обработкой и анализом данных, всё чаще включают несколько разнородных источников информации, объединенных в одну мультимодальную систему. Подтверждения этому можно найти в здравоохранении, аффективных вычислениях, робототехнических системах, образовании [1]. Разнородность данных приводит к необходимости в их эффективной обработке, а значит, совершенствованию методов представления, выравнивания, обобщении и генерации данных [1]. Таким образом, разработка подходов, способных справляться с многообразием структур и типов информации, становится ключевым направлением в области анализа данных и машинного обучения.

Мультимодальность - мощный инструмент для улучшения качества ответов модели [2]. Канонический корреляционный анализ (CCA) [3]

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 [4], [5] авторы показали, что он улучшает качество в задачах сопоставления событий. Однако CCA может моделировать лишь линейные зависимости.
Существует несколько подходов по улучшению CCA: Kernel-CCA, Correlaton Neural Network и Deep-CCA. В таблице (1) приведен небольшой анализ этих подходов.

Метод	**Особенность**
Kernel CCA	• стандартное расширение CCA на поиск нелинейных зависимостей • хорошая обработка высокоразмерных данных [6]; • скрытое представление ограничено фиксированным ядром; • является непараметрическим методом, что приводит к худшему масштабированию на новые данные [7].
Correlation Neural Network [4]	• подход на основе Autoencoders, который явно максимизирует корреляцию между представлениями при проецировании на общее подпространство.

Таблица 1: Сравнение подходов

В нашей работе мы рассмотрим улучшение CCA, связанное с механизмом Attention - нелинейным преобразованием последовательности для поиска сложных зависимостей. Так, усовершенствование cross-attention позволит лучше отсеивать информацию, снизит размерность пространства и тем самым повысит качество прогноза.
Ставя перед собой цель использовать преимущества обоих методов,

Метод	Особенность
Deep CCA [7]	• подход на основе Autoencoders, который явно максимизирует корреляцию между представлениями при проецировании на общее подпространство; • не считает скалярное произведение (inner product), что позволяет лучше масштабировать модель на новые данные.

Таблица 2: Сравнение подходов

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: победному, нейтральному или проигрышному удару.
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"[8].

3 Related Works

Интерфейс мозг-компьютер — это система, которая измеряет активность головного мозга и преобразует ее в (приблизительно) реальном времени в функционально полезные выходные сигналы для замены, восстановления, усиления, дополнения и/или улучшения естественных выходных мозговых сигналов, тем самым изменяя текущие процессы взаимодействия между мозгом и его внешней или внутренней средой [BCI Society]. Одним из наиболее распространенных неинвазивных методов получения информации об электрической активно-

сти мозга является ЭЭГ. Умение эффективно работать с этим типом данных полезно во многих задачах: распознавании эмоций [EEG-Based Emotion Recognition via Convolutional Transformer with Class Confusion-Aware Attention], прогнозировании рецедивов болезни [https://arxiv.org/abs/180 компенсации серьезных двигательных нарушений [https://iopscience.iop.org/arti 2552/aab2f2/pdf] и т.д.
написать про существующие attention cca подходы, см [https://eschola

4 Problem Statement

4.1 CCT: Canonical-Correlation Transformer

Мы продолжаем расширять область применимости CCA и представляем способ встраивания его в Attention.
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 афинные преобразования $\mathbf{W}_x, \mathbf{W}_y$, которые максимизируют корреляцию между X, Y в новом пространстве:

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

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 $C_{11} = \frac{1}{m} XX^\mathsf{T} \in \mathbb{R}^{n_1 \times n_1}$, $C_{22} = \frac{1}{m} YY^\mathsf{T} \in \mathbb{R}^{n_2 \times n_2}$, а матрица кросс-ковариации X, Y есть $C_{12} = \frac{1}{m} XY^\mathsf{T} \in \mathbb{R}^{n_1 \times n_2}$.

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

$$\begin{aligned}
\mathbf{W}_x^* &= C_{11}^{-\frac{1}{2}} U = (\frac{1}{m} X X^\mathsf{T})^{-\frac{1}{2}} U \\
\mathbf{W}_y^* &= C_{22}^{-\frac{1}{2}} V = (\frac{1}{m} Y Y^\mathsf{T})^{-\frac{1}{2}} V \\
\operatorname{corr}(\mathbf{W}_x^{\mathsf{T}*} X, \mathbf{W}_y^{\mathsf{T}*} Y) &= \operatorname{trace}(Z^\mathsf{T} Z)^{\frac{1}{2}}
\end{aligned} \tag{2}$$

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:

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

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.

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

$$\begin{aligned}
&\text{self-attn} : \mathbb{R}^{m \times n_1} \longrightarrow \mathbb{R}^{m \times d} \\
&\text{self-attn}(X) = \operatorname{attn}(XW_q, XW_k, XW_v)
\end{aligned} \tag{4}$$

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 $W^Q \in \mathbb{R}^{p \cdot d \times d}$:

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

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 $X_2 \in \mathbb{R}^{m \times d_2}$:

$$\text{cross-attn}(X_1, X_2) = \text{attn}(X_1 W_q, X_2 W_k, X_2 W_v) \tag{6}$$

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	$A = \frac{1}{\sqrt{d}} QK^\mathsf{T}$ – attention matrix	$\text{tr}(A^\mathsf{T} S_{12} B)$, s.t. $A^\mathsf{T} S_{11} A = B^\mathsf{T} S_{22} B = I$
Optimization Goal	Minimize task-specific loss	$\max_{A,B} \text{corr}(A^T X, B^T Y)$

Таблица 3: Comparison of Attention Mechanisms and CCA

Note that $A^\mathsf{T} S_{12} B = \frac{1}{m} A^\mathsf{T} X Y^\mathsf{T} B = \frac{1}{m} A^\mathsf{T} X \left(B^\mathsf{T} Y\right)^\mathsf{T} = \frac{1}{m} \widehat{Q} \widehat{K}^\mathsf{T}$. And it's quite similar to attention matrix formula $A = \dfrac{1}{\sqrt{d}} QK^\mathsf{T}$. 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	$W_Q^\mathsf{T} X$	$W_Q^\mathsf{T} X$	$A^\mathsf{T} X$	$S_{11}^{-\frac{1}{2}} X$	$S_{11}^{-\frac{1}{2}} X$
K	$W_K^\mathsf{T} X$	$W_K^\mathsf{T} Y$	$B^\mathsf{T} Y$	$S_{22}^{-\frac{1}{2}} Y$	$S_{22}^{-\frac{1}{2}} Y$
V	$W_V^\mathsf{T} X$	$W_V^\mathsf{T} Y$	I	$S_{11}^{-\frac{1}{2}} X$	$S_{22}^{-\frac{1}{2}} Y$
φ	softmax	softmax	Id	SVD_U	SVD_V

Таблица 4: United notation of CCA and attention

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

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

5 Experiments

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

5.1 Dataset Details

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

Мы оцениваем производительность модели на датасете "Real world table-tennis"[Dual-layer electroencephalography data during real-world table tennis]. Датасет включает в себя

5.2 Training Details

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

- На каких параметрах обучали сетку,

5.3 Experimental Results

Получилась вот такая точность и почему.

6 Appendix

6.1 CCA-Attention table

Приведем пример вывода значений таблицы для одного примера:

Список литературы

[1] Louis-Philippe Morency Paul Pu Liang, Amir Zadeh. Foundations and trends in multimodal machine learning: Principles, challenges, and open questions. *ACM Computing Surveys*, 2023.

[2] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. What makes multi-modal learning better than single (provably). In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 10944–10956. Curran Associates, Inc., 2021.

[3] Xinghao Yang, Weifeng Liu, Wei Liu, and Dacheng Tao. A survey on canonical correlation analysis. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2349–2368, 2021.

[4] Sarath Chandar, Mitesh M. Khapra, Hugo Larochelle, and Balaraman Ravindran. Correlational neural networks. *Neural Computation*, 28(2):257–285, 2016.

[5] Knani R. Hamdaoui F. et al. Bayoudh, K. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, 2022.

[6] Gretton Arthur Rauch Alexander Rainer Gregor Logothetis Nikos K. Müller Klaus-Robert Biebmann Felix, Meinecke Frank C. Temporal kernel cca and its application in multimodal neuronal data analysis. *Machine Learning*, 2010.

[7] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1247–1255, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[8] Amanda Studnicki and Daniel P. Ferris. Dual-layer electroencephalography data during real-world table tennis. *Data in Brief*, 52:110024, 2024.

[9] Nick Martin and Hermine Maes. Multivariate analysis. *London, UK: Academic*, 1979.

[10] 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.

[11] 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.

[12] 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.

[13] Yu Zhang Ehsan Adeli Qingyu Zhao Kilian M. Pohl Yixin Wang, Wei Peng. Brain-cognition fingerprinting via graph-gcca with contrastive learning. 2024.

[14] 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.

[15] 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.