ATD: Augmenting CP Tensor Decomposition by Self Supervision

NeurIPS 2022 | New Orleans
Nov 28th – Dec 9th

Chaoqi Yang[1], Cheng Qian[2], Navjot Singh[1], Cao Xiao[3], M Brandon Westover[4,5], Edgar Solomonik[1], Jimeng Sun[1]

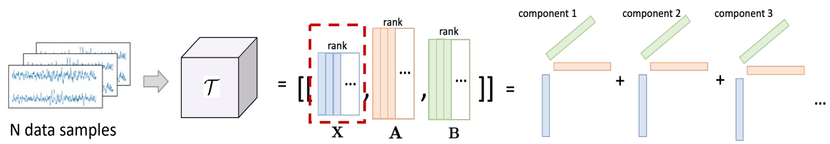
[1]University of Illinois Urbana-Champaign, [2]IQVIA, [3]Relativity, [4]Massachusetts General Hospital, [5]Harvard Medical School

I. Problem Definition

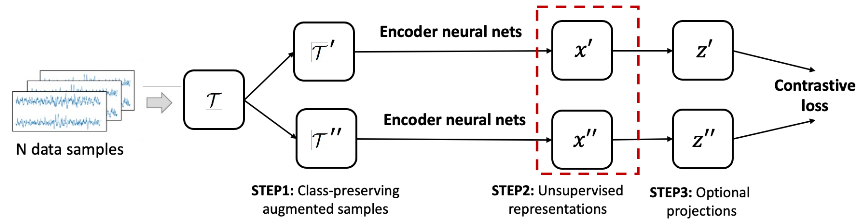
- Learn meaningful **unsupervised representations** for better **downstream classification.**

II. Unsupervised Feature Encoders

CP Tensor Decomposition (CPD) considers the **fitness property:**

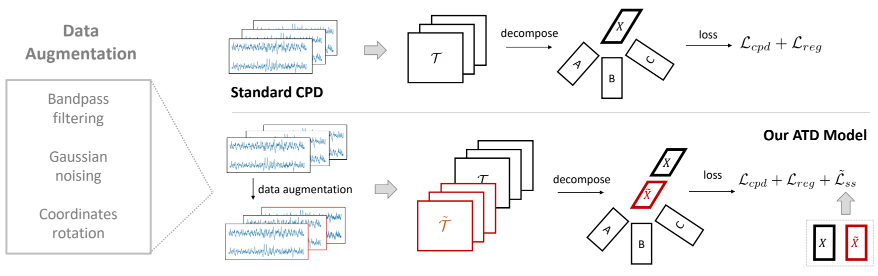


Self-supervised Contrastive Learning (SSL) considers the **alignment property**:



III. Our ATD Model: Augmenting CPD by SSL

- Our ATD model combines the **fitness** and the **alignment** properties with **regularizers**.



IV. Improved ALS for Non-convex Sub-iterations

- The overall objective function is convex to $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and **non-convex** to $\mathbf{X}, \tilde{\mathbf{X}}$
- We follow the alternating least squares (ALS) procedures to optimize one factor (e.g., $\mathbf{A}$) at a time
- Within each sub-iteration, we only use **fix-point iteration** and **least squares optimization**.

$$\mathcal{L} = \mathcal{L}_{cpd} + \mathcal{L}_{reg} + \tilde{\mathcal{L}}_{ss}$$

$$\mathcal{L}_{cpd} = \|\mathcal{T} - [\![\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|_F^2 + \left\|\tilde{\mathcal{T}} - [\![\tilde{\mathbf{X}}, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\right\|_F^2$$

$$\mathcal{L}_{reg} = \alpha\left(\|\mathbf{X}\|_F^2 + \|\tilde{\mathbf{X}}\|_F^2 + \|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2\right)$$

New self-supervised loss:

$$\tilde{\mathcal{L}}_{ss}^{\ominus}(\gamma) = (\gamma+1)\cdot\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{s\neq n}^{N}\left\langle\frac{\mathbf{x}^{(n)}}{\|\mathbf{x}^{(n)}\|_2}, \frac{\tilde{\mathbf{x}}^{(s)}}{\|\tilde{\mathbf{x}}^{(s)}\|_2}\right\rangle - \frac{1}{N}\sum_{n=1}^{N}\left\langle\frac{\mathbf{x}^{(n)}}{\|\mathbf{x}^{(n)}\|_2}, \frac{\tilde{\mathbf{x}}^{(n)}}{\|\tilde{\mathbf{x}}^{(n)}\|_2}\right\rangle$$

A 5-min Intro

Scan and understand our paper in 5 min

IV. Experimental Results

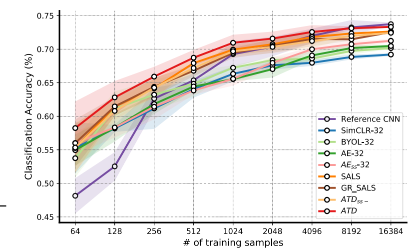
- **Datasets:** Sleep-EDF (EEG), Human activity recognition (HAR), PTB-XL (ECG), MGH (EEG)

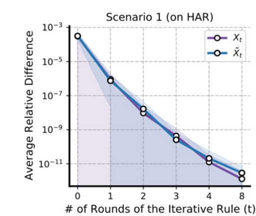
Name	Data Sample Format	Augmentations	# Unlabeled (N)	# Training	# Test	Task	# Class
Sleep-EDF	$I \times J \times K$: 14 × 129 × 86	(a), (b)	331,208	42,803	41,078	Sleep Staging	5
HAR	$I \times J \times K$: 18 × 33 × 33	(a), (b), (c)	7,352	1,473	1,474	Activity Recognition	6
PTB-XL	$I \times J \times K$: 24 × 129 × 75	(a), (b)	17,469	2,183	2,185	Gender Identification	2
MGH	$I \times J \times K$: 12 × 257 × 43	(a), (b)	4,377,170	238,312	248,041	Sleep Staging	5

Result 1: performance comparison on four datasets

	Sleep-EDF (5,000)		HAR (1,473)		PTB-XL (2,183)		MGH (5,000)	
	Accuracy	# of Params.	Accuracy	# of Params.	Accuracy	# of Params.	Accuracy	# of Params.
Self-sup models:								
SimCLR-32	84.98 ± 0.358	210,384	74.75 ± 0.723	53,286	69.25 ± 0.355	200,960	67.34 ± 0.970	212,624
SimCLR-128	**85.19 ± 0.358**	222,768	76.69 ± 0.697	65,670	68.19 ± 0.793	237,920	66.98 ± 1.331	246,608
BYOL-32	84.29 ± 0.405	211,440	73.71 ± 2.832	54,342	65.08 ± 1.535	202,016	68.83 ± 1.168	214,736
BYOL-128	83.26 ± 0.337	239,280	71.79 ± 1.866	82,182	65.49 ± 0.612	254,432	68.55 ± 1.339	279,632
Auto-encoders:								
AE-32	74.78 ± 0.723	217,216	63.13 ± 0.775	62,940	59.01 ± 0.896	224,528	68.58 ± 0.427	220,088
AE-128	75.17 ± 0.897	241,888	60.52 ± 1.604	87,612	58.29 ± 0.412	298,352	67.05 ± 1.375	257,048
AE_{ss}-32	80.92 ± 0.345	217,216	71.70 ± 2.135	62,940	68.47 ± 0.231	224,528	71.46 ± 0.386	220,088
AE_{ss}-128	81.84 ± 0.259	241,888	72.43 ± 1.370	87,612	68.88 ± 0.604	298,352	70.19 ± 0.617	257,048
Tensor models:								
SALS	84.27 ± 0.481	7,328	91.86 ± 0.295	2,688	69.15 ± 0.483	7,296	71.93 ± 0.379	9,984
GR-SALS	84.33 ± 0.356	7,328	92.33 ± 0.282	2,688	69.02 ± 0.477	7,296	72.35 ± 0.228	9,984
ATD_{ss-}	84.19 ± 0.221	7,328	92.41 ± 0.391	2,688	69.38 ± 0.612	7,296	72.78 ± 0.522	9,984
ATD	85.01 ± 0.224	7,328	**93.35 ± 0.357**	2,688	**70.26 ± 0.523**	7,296	**74.15 ± 0.431**	9,984

Result2: varying training data size





Result3: linear convergence of fix-point iteration

Acknowledgements

- This work was supported by NSF award SCH-2205289, SCH-2014438, IIS-1838042, NIH award R01 1R01NS107291-01.

Contacts

Email: chaoqiy2@illinois.edu
GitHub: github.com/ycq091044/ATD