Data Science

Selection of Tensor Representations For Multiview Forecasting

Student: *Nadezhda Alsahanova*
Skoltech Advisor: *Maxim Panov*

January, 2023

Skoltech

Area of research

Brain Computer Interfaces help to return communication and motor abilities.

Problem:
Initial data acquired by EEG or ECoG systems is redundant and highly correlated.

Source: GAO analysis (data). koya979/stock.adobe.com (images). | GAO-22-106118

Skoltech

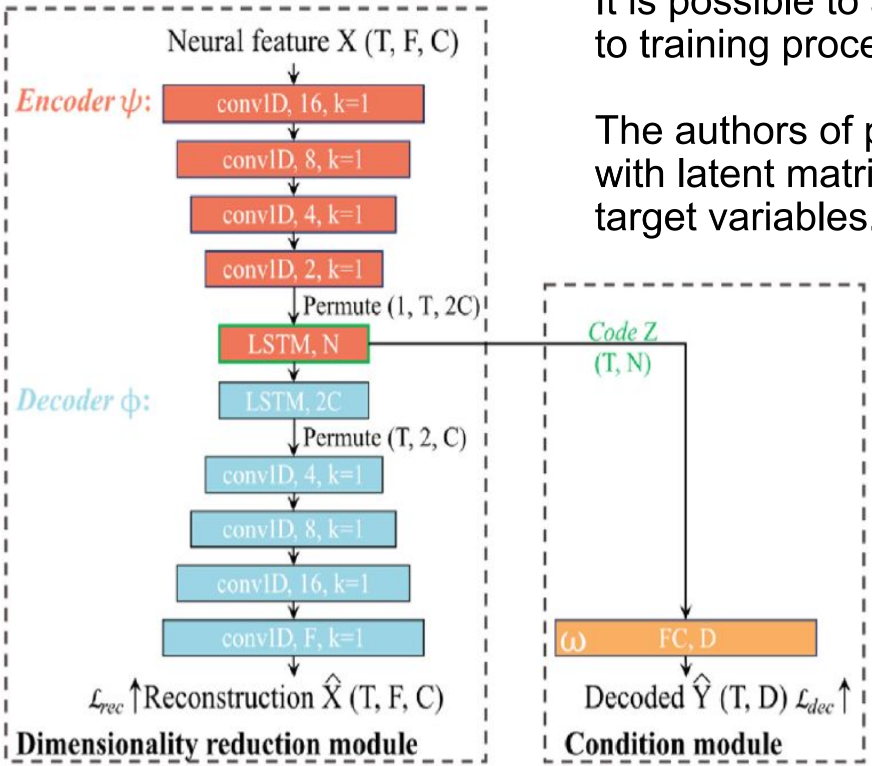
Current knowledge

<u>Solution</u>: dimensionality reduction.

Method	Nonlinear	Tensorial	Multiview
PCA	-	-	-
PLS, CCA	-	-	+
HOPLS	-	+	+
Autoencoders	+	+	-

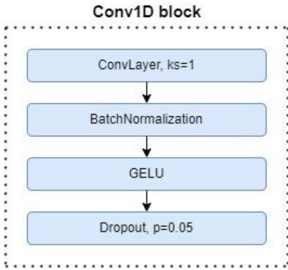
Skoltech

ReducedNet autoencoder



It is possible to add information about target variables to training procedure of Autoencoders.

The authors of paper [1] presented an autoencoder with latent matrix, that preserve information about target variables.



The authors used a loss:

$$L(\theta) = \mathscr{L}_{rec}(\theta_{AE}) + \lambda \mathscr{L}_{dec}(\theta_{CE})$$

Where:

$$\mathscr{L}_{rec}(\theta_{AE}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \psi\circ\varphi(X_i,\ \theta_{AE}))^2$$

$$\mathscr{L}_{dec}(\theta_{CE}) = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \omega\circ\varphi(X_i,\ \theta_{CE}))^2$$

<u>Problem</u>: Latent variable is matrix

[1] A hybrid autoencoder framework of dimensionality reduction for brain-computer interface decoding

Skoltech

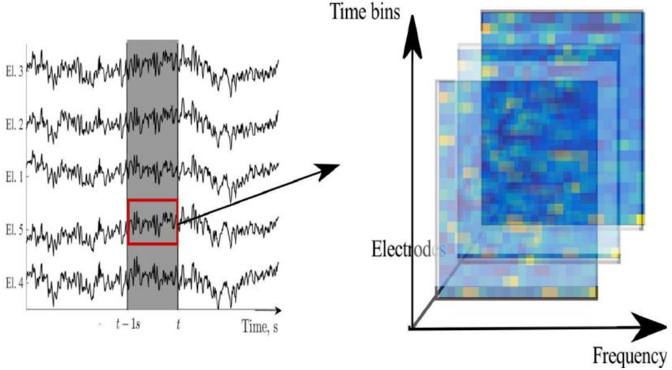
Tensorization

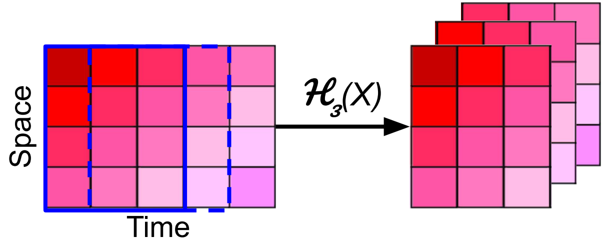
We can tensorize initial data to find low-rank approximation with high level of compression and to reveal hidden correlations.

There are several ways:

- Common one: <u>wavelet transform</u> of initial time series;

- Another one: <u>hankelization</u> of initial matrix or low-rank tensor by temporal or spatial dimensions.





Skoltech

Aim of thesis

The thesis goal is to find optimal representations of feature and target tensors in latent space by combining tensorization and dimensionality reduction methods. These tensor representations should be optimal in term of forecasting quality of the target variables and complexity of methods.

Skoltech

Objectives

- create tensor version of autoencoder and check whether it improves quality of decoding;

- check whether hankelizations along temporal, along spatial or along both modes improve quality of decoding;

- compare results of hankelization and autoencoder with state-of-the-art models.

Skoltech

Task of multiview forecasting

Let $\mathbf{s}(t),\ \mathbf{y}(t)$ are time series, where $\mathbf{y}(t)$ - target time series.

If there are several sources of initial time series $\mathbf{s}(t)$, the dataset made from these time series:

$$\underline{\mathbf{X}} \in \mathbb{R}^{T \times n_1 \times \ldots \times n_D}, \qquad \mathbf{Y} \in \mathbb{R}^{T \times K}$$

Partial Least Squares (PLS)

PLS works with data presented in **matrices**. It projects dependent and independent variables into latent space:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^\top + \mathbf{F} = \sum_{k=1}^{\ell} \mathbf{t}_k \cdot \mathbf{p}_k^\top + \mathbf{F}$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^\top + \mathbf{E} = \sum_{k=1}^{\ell} \mathbf{u}_k \cdot \mathbf{q}_k^\top + \mathbf{E}.$$

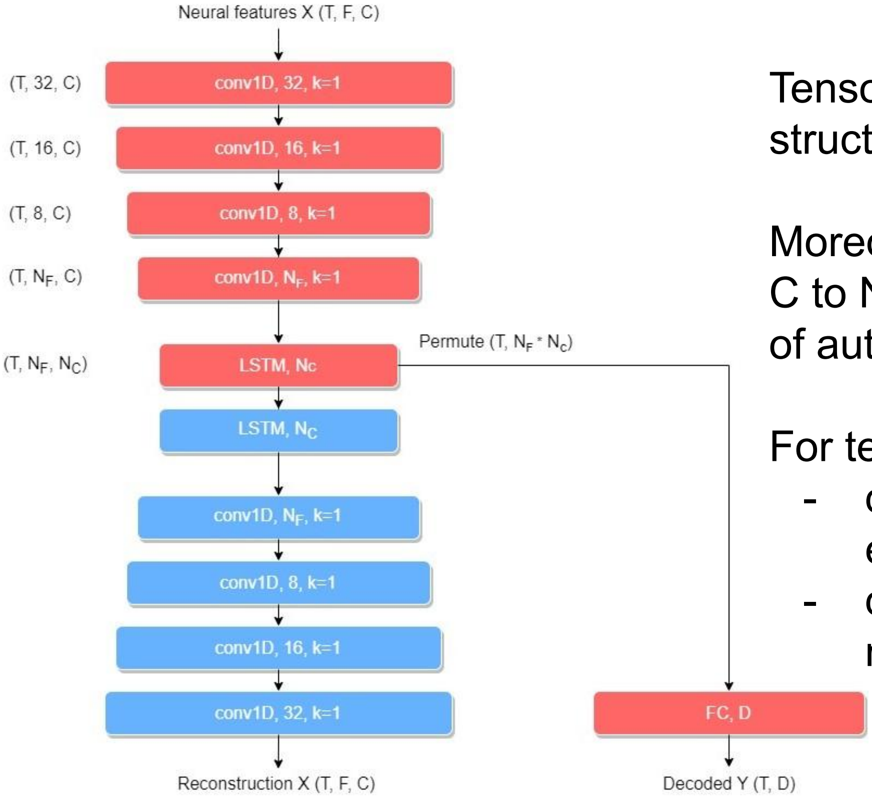
Where:

$\mathbf{T}, \mathbf{U}$ are the latent space matrices; $\mathbf{P}, \mathbf{Q}$ are the loading matrices.

So that:

$$cov(\mathbf{Xp}, \mathbf{Yq})^2 = \frac{\mathbf{p}^\top \mathbf{X}^\top \mathbf{Y} \mathbf{q}}{\|\mathbf{p}\|_2 \|\mathbf{q}\|_2} \to \max_{\mathbf{p},\ \mathbf{q}}$$

TensorReducedNet



TensorReducedNet allows not to lose tensor structure of initial data.

Moreover it uses LSTM layer only to reduce from C to N_C, and that allows to have less parameters of autoencoder.

For tensors with order higher than 3:

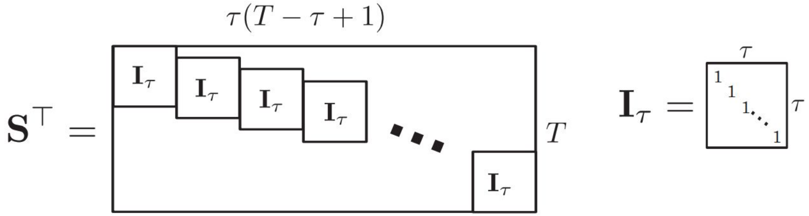
- decrease of dimensionality of all modes except one made by convolutional blocks
- decrease of dimensionality of the last mode made by LSTM.

Skoltech

Hankelization

Hankelization is an effective way to transform lower-order data to higher-order tensors.

Hankelization of tensor $\underline{\mathbf{X}} \in \mathbb{R}^{T \times n_1 \times \ldots, n_D}$ can be done by multi-way delay embedding transform (MDT) with use of matrix $\mathbf{S}$:



$$\hat{\underline{\mathbf{X}}} = \mathcal{H}_\tau\left(\underline{\mathbf{X}}\right) = Fold_{(T,\tau)}\left(\underline{\mathbf{X}} \times_1 \mathbf{S}\right) \in \mathbb{R}^{(T-\tau+1)\times\tau\times n_1 \times \ldots \times n_D}$$

Where $Fold_{(T,\tau)} : \mathbb{R}^{\tau(T-\tau+1)\times n_1 \times \ldots \times n_D} \to \mathbb{R}^{(T-\tau+1)\times\tau\times n_1 \times \ldots \times n_D}$. The inverse MDT:

$$\underline{\mathbf{X}} = \mathcal{H}_\tau^{-1}\left(\hat{\underline{\mathbf{X}}}\right) = Unfold_{(T,\tau)}\left(\hat{\underline{\mathbf{X}}}\right) \times_1 \mathbf{S}^\dagger$$

It was a hankelization along the temporal mode. Similarly, it can be done for any spatial modes.

Skoltech

Tensor Regression

We are trying to avoid matricization on every step of decoding target variables. So, tensor regression can be defined as:

$$\mathbf{y}_m = \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle + \varepsilon$$

Where $\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle$ denotes a tensor contraction along the first D modes:

$$\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle_k = \sum_{i_1=1}^{n_1} \cdots \sum_{i_D=1}^{n_D} x_{i_1,\dots,i_D} w_{i_1,\dots,i_D,k}$$

In practice, for very large scale problems, tensors are expressed approximately in tensor network formats. For example, with the application of Tucker multilinear rank tensor representation:

$$\underline{\mathbf{W}} \approx \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \dots \times_D \mathbf{U}^{(D)} \times_{D+1} \mathbf{U}^{(D+1)}$$

Skoltech

Algorithm

$$\begin{tikzcd}
\underline{\underline{\mathbf{X}}} \arrow[r, "\Phi"] \arrow[d, "\Psi_x"'] & \mathbf{Y} \arrow[d, "\Psi_y"'] \\
\underline{\underline{\hat{\mathbf{X}}}} \arrow[r, "\underline{\Phi}"] \arrow[d, "h_x"'] & \underline{\mathbf{Y}} \arrow[d, "h_y"'] \\
\mathbf{T} \arrow[r, "g"'] & \mathbf{U}
\end{tikzcd}$$

$$\Phi = \Psi_x \circ h_x \circ g \circ h_y^{-1} \circ \Psi_y^{-1}$$

Ψ_x, Ψ_y are tensorization methods

h_x ,h_y are dimensionality reduction models

g is regression model in latent space

$\underline{\Phi}$ can be presented by PLS, HOPLS or CCA

$\mathbf{T}, \mathbf{U}$ are latent tensors

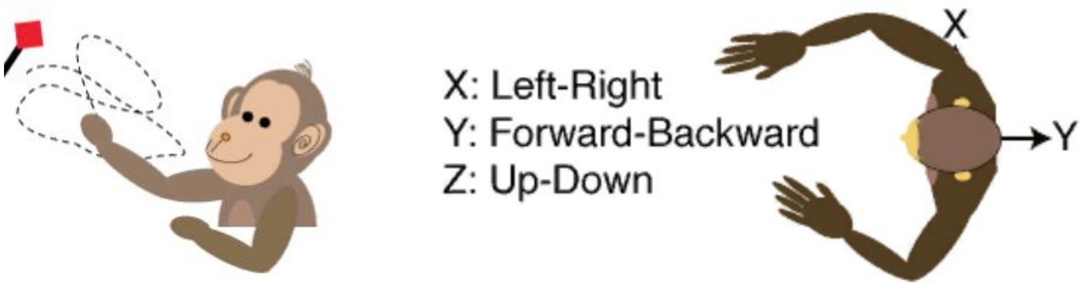
NeuroTycho foodtracking dataset

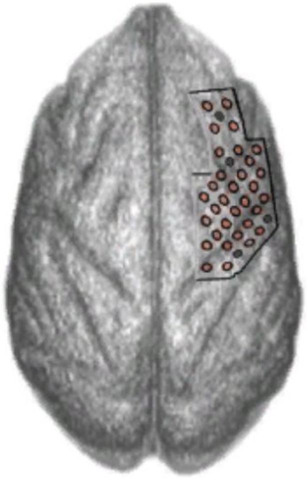
ECoG signals was obtained from 32 channels. Moreover, frequency-domain features are obtained with wavelet transform with 27 frequencies.

$$\underline{\mathbf{X}} \in \mathbb{R}^{T\times 32\times 27}, \ \mathbf{Y} \in \mathbb{R}^{T\times 3}$$

After hankelization along temporal and spatial modes:

$$\hat{\underline{\mathbf{X}}} \in \mathbb{R}^{\hat{T}\times 10\times 27\times 31\times 2}, \underline{\mathbf{Y}} \in \mathbb{R}^{\hat{T}\times 10\times 3}$$





Results

$$\mathrm{sRMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_{\mathbf{a}}) = \sqrt{\frac{\mathrm{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_{\mathbf{a}})}{\mathrm{MSE}(\mathbf{Y}, \overline{\mathbf{Y}})}} = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}_{\mathbf{a}}\|_2}{\|\mathbf{Y} - \overline{\mathbf{Y}}\|_2}$$

, where $\widehat{\mathbf{Y}}_{\mathbf{a}}$ - prediction.

Model	Hankelization	sRMSE	Shape of latent variable	Number of parameters of NN
PLS	-	0.978	4	
ReducedNet	-	0.971	2	420792
TensorReducedNet	-	0.965	11 x 5	9446
TensorReducedNet	+	**0.928**	3 x 3 x 3 x 2	**7508**

Skoltech

Plans

- Tune <u>parameters of hankelization</u>
- Check influence of <u>hankelization along spatial dimension</u>
- Reproduce the results on <u>another BCI dataset</u>
- <u>Compare results</u> with more SOTA models

Skoltech

Thank you for your attention!