

# Выбор тензорных представлений для прогнозирования по мультимодальным измерениям.

Алсаханова Надежда

Московский физико-технический институт  
Кафедра интеллектуальных систем

Научный руководитель д.ф.-м.н. В. В. Стрижов

Москва, 2022

# Данные

## Определение

*Временным рядом* называется функция дискретного аргумента  $\mathbf{s}(t)$ , сопоставляющая отчетам времени  $t_i \in \mathcal{T}$  вектор значения измеряемых переменных  $\mathbf{s}(t_i) = \mathbf{s}_i \in \mathbb{R}^M$ .

## Определение

Пусть заданы набор временных рядов  $\{\mathbf{s}_n(t)\}_{n=1}^{N_s}$  и набор целевых временных рядов  $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$ . Задача восстановления значений  $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$  по предыдущим значениям  $\{\mathbf{s}_n(t)\}_{n=1}^{N_s}$  называется задачей декодирования временных рядов  $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$ .

Пусть из временных рядов  $\mathbf{s}(t)$ ,  $\mathbf{y}(t)$  составлена выборка  $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ :

$$\underline{\mathbf{X}} \in \mathbb{R}^{M \times n_1 \times \dots \times n_D}, \quad \underline{\mathbf{Y}} \in \mathbb{R}^{M \times K}, \quad (1)$$

где  $\mathbf{y}_m = \mathbf{y}(t_m)$ , а  $\underline{\mathbf{X}}_m \in \mathbb{R}^{n_1 \times \dots \times n_D}$  - тензор.

# Постановка задачи

## Определение

**Модель снижения размерности**  $h : \mathbb{R}^{n_1 \times \dots \times n_D} \rightarrow \mathbb{R}^{\ell_1 \times \dots \times \ell_D}$ , где  $\ell_i < n_i$  является преобразованием, которое переводит из пространства большей размерности  $\mathbb{R}^{n_1 \times \dots \times n_D}$  в пространство меньшей размерности  $\mathbb{R}^{\ell_1 \times \dots \times \ell_D}$ , то есть  $\underline{\mathbf{Q}}_m = h(\underline{\mathbf{X}}_m, \theta)$ , где  $\theta$  – параметры преобразования.

## Определение

**Модель согласования**  $g : \mathbb{R}^{\ell_1 \times \dots \times \ell_D} \rightarrow \mathbb{R}^K$  – преобразование, переводящая из пространства меньшей размерности  $\mathbb{R}^{\ell_1 \times \dots \times \ell_D}$  в пространство целевой переменной  $\mathbb{R}^K$ , то есть  $\mathbf{Y}_m = g(\underline{\mathbf{Q}}_m, \eta)$ , где  $\eta$  – параметры преобразования.

# Постановка задачи

Предлагаемая модель:

$$\phi = h \circ g : \mathbb{R}^{n_1 \times \dots \times n_D} \rightarrow \mathbb{R}^K, \quad (2)$$

где функции  $h$  и  $g$  удовлетворяют процессу:

$$\underline{\mathbf{X}}_m \xrightarrow{h(\cdot, \theta)} \underline{\mathbf{Q}}_m \xrightarrow{g(\cdot, \eta)} \mathbf{Y}_m$$

Модель  $\phi^*(\cdot, \theta, \eta)$  является оптимальной, если она минимизирует некоторый функционал ошибки  $\mathcal{L}$ :

$$\phi^* = \arg \min_{\{\theta, \eta\}} \mathcal{L}(\phi(\underline{\mathbf{X}}, \theta, \eta), \mathbf{Y}) \quad (3)$$

# Мультимодальная регрессия

Тензор признаков  $\underline{\mathbf{X}}$  можно матрицизовать  $\underline{\mathbf{X}}$  по первой размерности:

$$\underline{\mathbf{X}}_{(1)} = \left[ \text{vec}(\underline{\mathbf{X}}_1)^\top, \dots, \text{vec}(\underline{\mathbf{X}}_M)^\top \right]^\top \in \mathbb{R}^{M \times (n_1 \cdots n_D)} \quad (4)$$

Тогда задача сводится к задаче мультимодальной регрессии:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \mathcal{L}(\mathbf{f}, \underline{\mathbf{X}}_{(1)}, \mathbf{Y}) \quad (5)$$

Рассматриваем только класс параметрических моделей  $\mathbf{f}(\mathbf{X}, \Theta)$ , где  $\Theta$  – матрица параметров. Тогда оптимальные параметры:

$$\Theta^* = \arg \min_{\Theta} \mathcal{L}(\Theta, \mathbf{X}, \mathbf{Y}) \quad (6)$$

Избыточность матрицы  $\underline{\mathbf{X}}_{(1)}$  приводит к неустойчивости решения (6).

# Тензорная регрессия

Матрицизация тензора  $\underline{\mathbf{X}}$  плоха тем, что:

- получается чрезмерно большое количество столбцов;
- теряется информация для структурированных данных.

Тензорная регрессия:

$$\mathbf{y}_m = \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle + \varepsilon, \quad (7)$$

где  $\varepsilon \in \mathbb{R}^K$  – ошибка,  $\mathbf{y}_m \in \mathbb{R}^K$  – целевой вектор, а  $\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle$  – тензорная свертка:

$$\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle_k = \sum_{i_1=1}^{n_1} \cdots \sum_{i_D=1}^{n_D} x_{i_1, \dots, i_D} w_{i_1, \dots, i_D, k} \quad (8)$$

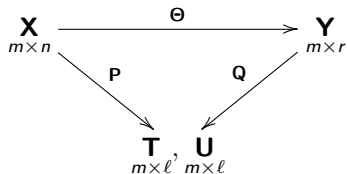
Тензор  $\underline{\mathbf{W}}$  представляют в разложении Такера:

$$\underline{\mathbf{W}} \approx \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \cdots \times_D \mathbf{U}^{(D)} \times_{D+1} \mathbf{U}^{(D+1)} \quad (9)$$

где  $\underline{\mathbf{G}}$  – центральный тензор меньшего размера,  $\mathbf{U}^{(i)}$  – унитарные матрицы.

# Partial Least Squares, PLS

Задана выборка  $(\mathbf{X}, \mathbf{Y})$ , где  $\mathbf{X} \in \mathbb{R}^{m \times n}$  — матрицы объектов,  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  — матрица ответов.



- максимизировать ковариацию между образцами;
- сохранить информацию об исходных матрицах.

$$\mathbf{X}_{m \times n} = \mathbf{T}_{m \times \ell} \cdot \mathbf{P}_{\ell \times n}^T + \mathbf{F}_{m \times n} = \sum_{k=1}^{\ell} \mathbf{t}_k_{m \times 1} \cdot \mathbf{p}_k^T_{1 \times n} + \mathbf{F}_{m \times n},$$

$$\mathbf{Y}_{m \times r} = \mathbf{U}_{m \times \ell} \cdot \mathbf{Q}_{\ell \times r}^T + \mathbf{E}_{m \times r} = \sum_{k=1}^{\ell} \mathbf{u}_k_{m \times 1} \cdot \mathbf{q}_k^T_{1 \times r} + \mathbf{E}_{m \times r}.$$

# Higher-Order Partial Least Squares, HOPLS

Задана выборка  $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ , где  $\underline{\mathbf{X}} \in \mathbb{R}^{M \times I_1 \times \dots \times I_N}$  — тензор объектов,  $\underline{\mathbf{Y}} \in \mathbb{R}^{M \times J_1 \times \dots \times J_N}$  — тензор ответов.

Для тензоров  $\underline{\mathbf{X}}$  и  $\underline{\mathbf{Y}}$  используется разложение Такера:

$$\underline{\mathbf{X}} = \sum_{r=1}^R \underline{\mathbf{G}}_{xr} \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \times_3 \dots \times_{N+1} \mathbf{P}_r^{(N)} + \underline{\mathbf{E}}_R,$$
$$\underline{\mathbf{Y}} = \sum_{r=1}^R \underline{\mathbf{G}}_{yr} \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \times_3 \dots \times_{N+1} \mathbf{Q}_r^{(N)} + \underline{\mathbf{F}}_R,$$

где  $R$  — количество векторов в скрытом пространстве,

- $\mathbf{t}_r \in \mathbb{R}^M$  —  $r$ -ый скрытый вектор;
- $\{\mathbf{P}_r^{(n)}\}_{n=1}^N \in \mathbb{R}^{I_n \times L_n}$ ,  $\{\mathbf{Q}_r^{(n)}\}_{n=1}^N \in \mathbb{R}^{J_n \times K_n}$  — факторные матрицы;
- $\underline{\mathbf{G}}_{xr} \in \mathbb{R}^{1 \times L_1 \times \dots \times L_N}$ ,  $\underline{\mathbf{G}}_{yr} \in \mathbb{R}^{1 \times K_1 \times \dots \times K_N}$  — ядра разложения Такера.



# Higher-Order Partial Least Squares, HOPLS

Перепишем разложение Такера следующим образом:

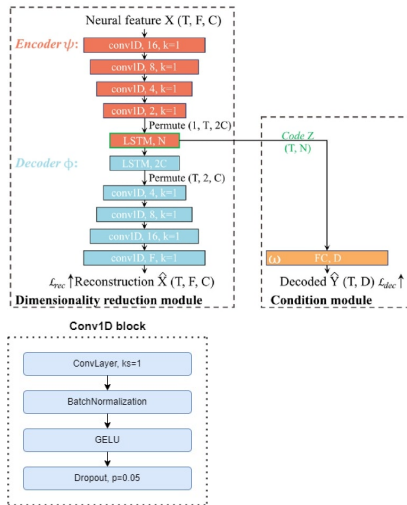
$$\begin{aligned}\underline{\mathbf{X}} &= \underline{\mathbf{G}}_x \times_1 \mathbf{T} \times_2 \overline{\mathbf{P}}^{(1)} \dots \times_{N+1} \overline{\mathbf{P}}^{(N)} + \underline{\mathbf{E}}_R, \\ \underline{\mathbf{Y}} &= \underline{\mathbf{G}}_y \times_1 \mathbf{T} \times_2 \overline{\mathbf{Q}}^{(1)} \dots \times_{N+1} \overline{\mathbf{Q}}^{(N)} + \underline{\mathbf{F}}_R,\end{aligned}$$

где

- $\underline{\mathbf{E}}_R$ ,  $\underline{\mathbf{F}}_R$  – тензоры остатков, полученные после извлечения  $R$  скрытых компонентов;
- $\underline{\mathbf{G}}_x$ ,  $\underline{\mathbf{G}}_y$  – ядра разложения, которые имеют блочно-диагональную структуру.

Элементы  $\underline{\mathbf{G}}_x$ ,  $\underline{\mathbf{G}}_y$  указывают на уровень локальных взаимодействий между соответствующими скрытыми векторами и факторными матрицами.

# Автоэнкодер ReducedNet



Автоэнкодер ReducedNet был применен для задачи с данными ECoG. Тензор признаков –  $\mathbf{X} \in \mathbb{R}^{T \times F \times C}$ , где  $F$  – количество частот,  $C$  – количество каналов.

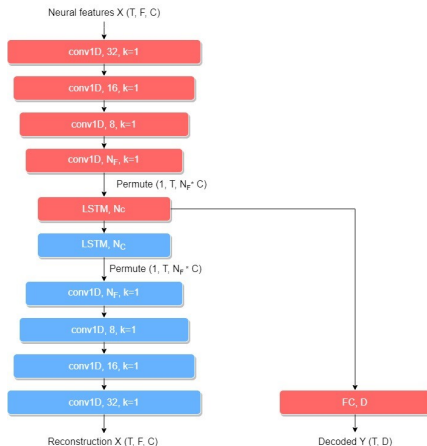
Функция потерь для обучения автоэнкодера:

$$\mathcal{L}_1 = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{dec}, \quad (10)$$

$$\mathcal{L}_{rec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{x}_m - h \circ \psi(\mathbf{x}_m)\|^2$$

$$\mathcal{L}_{dec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{y}_m - \omega \circ h(\mathbf{x}_m)\|^2$$

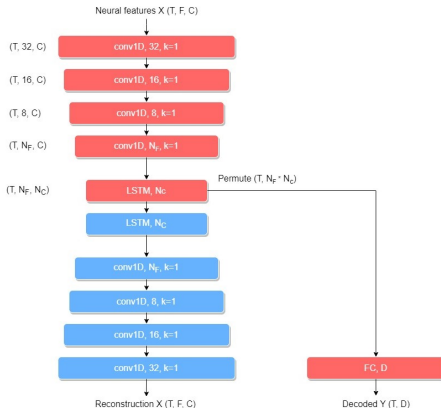
# Автоэнкодер MatrixReduceNet



Отличия:

- Введен новый параметр сети  $N_F$ ;
- Увеличено количество каналов.

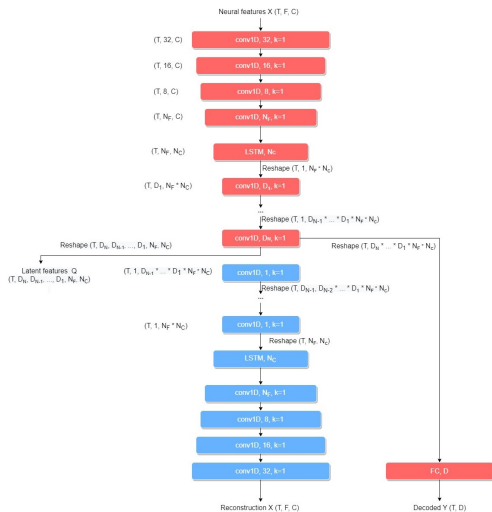
# Автоэнкодер TensorReduceNet



Отличия:

- Не теряется исходная структура данных, то есть на выходе энкодера тензор.

# Автоэнкодер TensorDimReduceNet



Можно создавать на выходе из энкодера тензора с дополнительными размерностями (дополнительные нелинейные признаки).

## Другая функция потерь

Добавляется к функции потерь из статьи компонент с корреляций столбцов в тензоре на выходе из энкодера:

$$\mathcal{L}_2 = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{dec} + \beta \cdot \mathcal{L}_{cor}, \quad (11)$$

где:

$$\mathcal{L}_{rec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{X}_m - h \circ \psi(\mathbf{X}_m)\|^2$$

$$\mathcal{L}_{dec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{Y}_m - \omega \circ h(\mathbf{X}_m)\|^2$$

$$\mathcal{L}_{cor} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{corr}(h(\mathbf{X}_m)_i, h(\mathbf{X}_m)_j)$$

где  $n$  - количество столбцов в тензоре  $h(\mathbf{X}_m)$ .

# Сравнение

Модель	Учитывает	Нелинейность	Работа с тензорами
PCA	—	—	—
PLS	+	—	—
HOPLS	+	—	+
Autoencoder	—	+	+
Our autoencoder	+	+	+

# Датасет Sensors

$X \in \mathbb{R}^{T \times 6 \times 15}$ , где 15 – количество частот,  $Y \in \mathbb{R}^{T \times 3}$ .

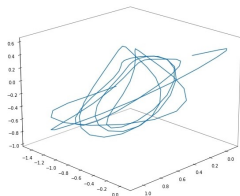


Рис.: Данные с гироскопа

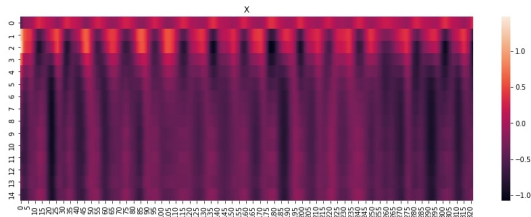


Рис.: Признаки в частотно-временном пространстве



# Датасет ECoG

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

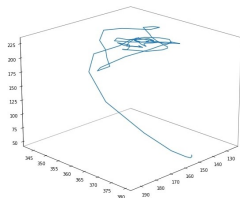


Рис.: Траектория движения руки

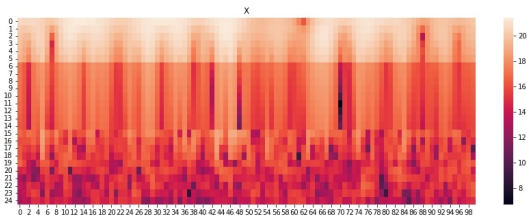
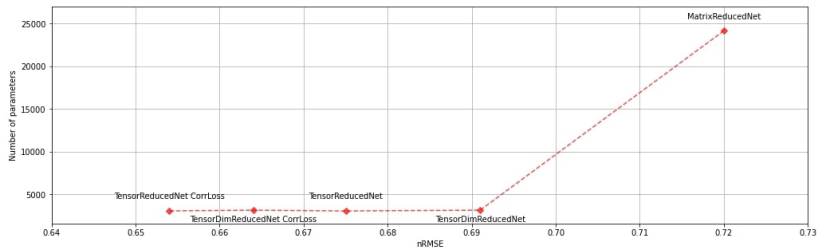


Рис.: Данные с одного канала в частотно-временном пространстве

## Результаты на Sensors

Модель	Компоненты	RMSE	Параметры
PLS	5	0.757	-
HOPLS	$3 \times 1$	0.811	-
MatrixReducedNet	6	0.720	24163
TensorReducedNet+LR	$5 \times 2$	0.682	3069
TensorReducedNet+TR	$5 \times 2$	<b>0.675</b>	<b>3069</b>
TensorDimReducedNet+LR	$2 \times 6 \times 2$	0.710	3163
TensorDimReducedNet+TR	$2 \times 6 \times 2$	0.691	3163
TensorReducedNet+LR (corr)	$2 \times 3$	0.705	3076
TensorReducedNet+TR (corr)	<b><math>2 \times 3</math></b>	<b>0.654</b>	<b>3076</b>
TensorDimReducedNet+LR (corr)	$6 \times 4 \times 1$	0.704	3165
TensorDimReducedNet+TR (corr)	$6 \times 4 \times 1$	0.664	3165

# Результаты на Sensors



## Результаты на ECoG

Модель	Компоненты	RMSE	Параметры
PLS	<b>4</b>	0.957	-
HOPLS	<b>1 × 1</b>	1.138	-
MatrixReducedNet	19	0.925	57106
TensorReducedNet+LR	6 × 14	0.926	12418
TensorReducedNet+TR	6 × 14	<b>0.922</b>	<b>12418</b>
TensorDimReducedNet+LR	2 × 5 × 14	0.937	12808
TensorDimReducedNet+TR	2 × 5 × 14	0.931	12808
TensorReducedNet+LR (corr)	<b>6 × 9</b>	<b>0.939</b>	<b>10548</b>
TensorReducedNet+TR (corr)	6 × 9	0.941	10548
TensorDimReducedNet+LR (corr)	2 × 4 × 10	0.970	11169
TensorDimReducedNet+TR (corr)	2 × 4 × 10	0.973	11169

# Результаты на ECoG

