

Тензорная декомпозиция и прогноз для набора временных рядов

Сёмкин Кирилл

Московский физико-технический институт
Кафедра интеллектуальных систем

Научный руководитель: д.ф.-м.н. Стрижов Вадим Викторович
2024

Мотивация

Проблема

Обработка многомерных временных рядов влечёт необходимость учёта взаимосвязей сигналов. Нейросетевые и статистические подходы имеют большое количество параметров, и требуют изощрённых процедур обучения. При этом они не предлагают способы построения декомпозиции.

Цель

Предложить метод, позволяющий выделить общую для набора сигналов структуру. На её основании произвести разложение каждого сигнала на компоненты. Найти способ построения прогноза.

Решение

tSSA = модель собственного пространства сигнала + тензорное представление данных + CPD

Модель порождения наблюдаемых сигналов

Динамическая система в дискретном времени:

$$\begin{cases} \mathbf{y}(t+1) = f(\mathbf{y}(t)), & t \in \mathbb{N} \\ \mathbf{y}(0) = \mathbf{y}_0 \end{cases}$$

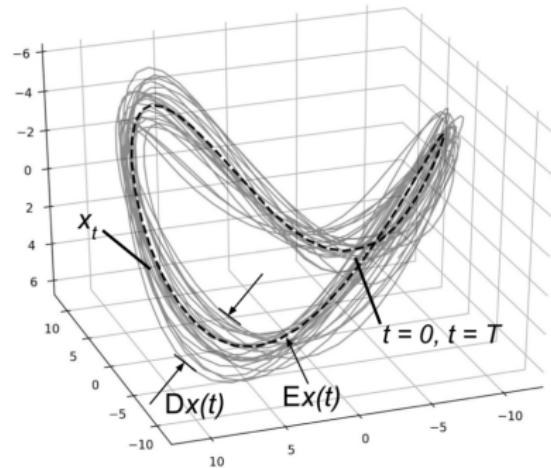
$\mathbf{y} \in X$, где X — гладкое многообразие большой размерности.

Конкретная траектория этой системы порождает наблюдаемые ряды через некое многомерное отображение $\varphi : X \rightarrow \mathbb{R}^m$:

$$\varphi(\mathbf{y}(t)) = \mathbf{x}_t \Leftrightarrow \begin{cases} \varphi_1(\mathbf{y}(t)) = x_1(t) \\ \dots \\ \varphi_m(\mathbf{y}(t)) = x_m(t) \end{cases}$$

Низкоразмерное представление динамической системы

Траектории $y(t)$ лежат в многообразии $M \subset X$ размерности меньшей, чем у X . Ставится задача поиска вложения (embedding) M в \mathbb{R}^L для некоторого L и нахождения базиса в образе этого вложения.



Динамическая система собственной размерности < 3

Решение для одного сигнала

Theorem (Такенс)

Искомое вложение даётся построением соответствующего вектора задержки

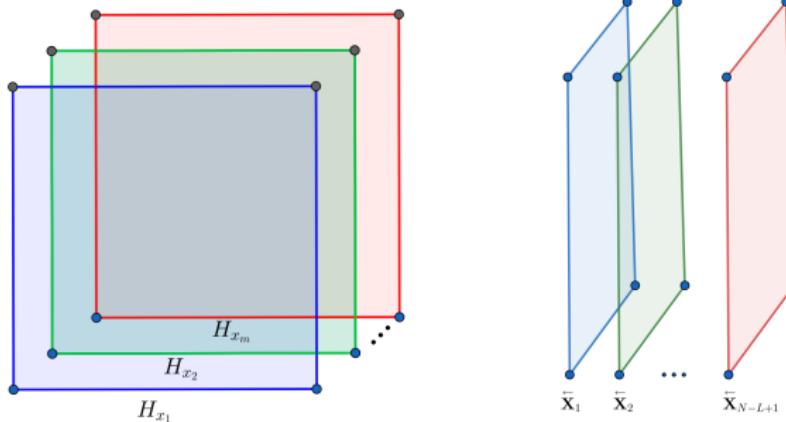
$$y(t) \xrightarrow{\text{emb}} \begin{pmatrix} \varphi \circ f^{t-L+1}(y(t)) \\ \vdots \\ \varphi \circ f(y(t)) \\ \varphi \circ y(t) \end{pmatrix} = \begin{pmatrix} x(t-L+1) \\ \vdots \\ x(t-1) \\ x(t) \end{pmatrix} = \overleftarrow{x}_t \in \mathbb{R}^L$$

Полученное пространство вложения $\text{Lin}(\{\overleftarrow{x}_t\})$ есть собственное пространство сигнала $x(t)$.

С помощью разложения траекторной матрицы $H_x = [\overleftarrow{x}_1 \dots \overleftarrow{x}_{N-L+1}]$. выделяем базис. Далее можем декомпозировать и строить прогноз (SSA).

Построение траекторного тензора

Упакуем векторы задержек всех рядов в матрицы задержек
 $(\overleftarrow{\mathbf{x}}_{1_t} \dots \overleftarrow{\mathbf{x}}_{m_t}) := \overleftarrow{\mathbf{X}}_t$. Состыкуем их в траекторный тензор \mathbf{T} .



Два вида на \mathbf{T} . Слева — в виде набора траекторных матриц сигналов.
Справа — в виде набора матриц задержки.

Решение для нескольких сигналов

Применим *CP-разложение* к траекторному тензору и рассмотрим каждое его сечение по третьему измерению:

$$\mathbf{T} = \sum_{i=1}^r \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \Leftrightarrow \begin{cases} \mathbf{H}_{x_1} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_1}(i) \cdot \mathbf{a}_i \mathbf{b}_i^T \\ \mathbf{H}_{x_2} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_2}(i) \cdot \mathbf{a}_i \mathbf{b}_i^T \\ \dots \\ \mathbf{H}_{x_m} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_m}(i) \cdot \mathbf{a}_i \mathbf{b}_i^T \end{cases}$$

Получили разложение траекторных матриц сигналов по общему базису $\{\mathbf{a}_i\}_{i=1}^r$. Связь рядов определяют $\boldsymbol{\sigma}_{x_k}$, имеющие смысл сингулярных чисел.

Аддитивное разложение сигналов

Принцип построения: декомпозиция траекторной матрицы H_{x_k} определяет декомпозицию временного ряда $x_k(t)$.

Факторы разложения H_{x_k} разбиваем на группы и ганкелизуем — усредняем матрицы по антидиагоналям.

$$\begin{aligned} H_{x_k} &= \sum_{i=1}^r \sigma_{x_k}(i) \cdot a_i b_i^T = \sum_{i \in \mathbb{I}_1} \sigma_{x_k}(i) \cdot a_i b_i^T + \dots + \sum_{i \in \mathbb{I}_s} \sigma_{x_k}(i) \cdot a_i b_i^T = \\ &= C_1 + \dots + C_s = \text{Hankel}(C_1) + \dots + \text{Hankel}(C_s) \Leftrightarrow x_k(t) = c_1(t) + \dots + c_s(t) \end{aligned}$$

Наблюдение

Чем ближе по структуре C_i похожи на ганкелевы, тем более естественно построение итогового разложения.

Оптимальная декомпозиция

Обозначим невязку $H_i = \sigma_{x_k}(i) \cdot a_i b_i^T - \text{Hankel}(\sigma_{x_k}(i) \cdot a_i b_i^T)$. Тогда из предыдущего выражения следует:

$$H_1 + \dots + H_r = 0$$

Разбиваем на две группы. В каждой хотя бы две невязки, суммирующиеся в ноль. Условие оптимальной декомпозиции:

$$\begin{cases} \sum_{j=1}^{r-1} \beta_j H_j = 0 \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_j \geq 2 \end{cases} \rightarrow \begin{cases} \|\Lambda \beta\| \rightarrow \min_{\beta} \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_j \geq 2 \end{cases}$$

Имеем задачу наименьших квадратов с целочисленными ограничениями (ILS), которая является NP-сложной.

Построение среднего прогноза

Базис в пространстве векторов задержек сигнала даётся как
 $\text{Lin}(\{a_i\}) \Leftrightarrow A = [a_1 \dots a_r]$.

Прогноз на один шаг вперёд сводится к решению частично неизвестной СЛАУ:

$$\overleftarrow{x}_{N+1} = A\lambda \Leftrightarrow \begin{cases} x_{kn} = A_{kn}\lambda \\ x(N+1) = a_{pr}^T \lambda \end{cases}, \text{ где}$$

$$A = \begin{pmatrix} A_{kn} \\ a_{pr}^T \end{pmatrix}$$

$$\overleftarrow{x}_{N+1} = (x_{kn}, x(N+1))^T$$

Система переопределена, решение даётся в смысле наименьших квадратов:

$$x(N+1) = a_{pr}^T (A_{kn}^T A_{kn})^{-1} A_{kn}^T x_{kn}$$

Постановка эксперимента

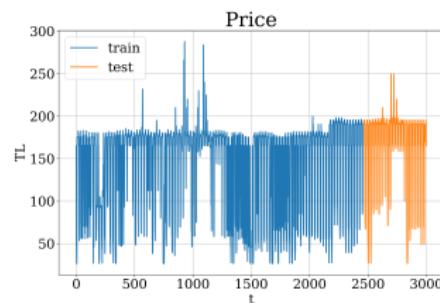
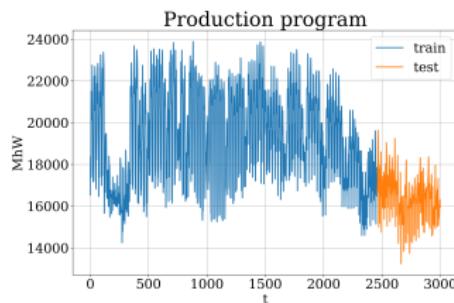
Рассматриваемые ряды:

- 1 план выработки электричества на день и цена его производства
- 2 показания акселерометра и гироскопа (ходьба)

Модели для сравнения: mSSA, VAR, RNN.

Метрики:

- 1 Прогнозирование: MSE, MAPE
- 2 Декомпозиция: AHE, RHE



Потребление электричества и его цены за единицу мощности

Метрики декомпозиции

Определение

Абсолютной ошибкой ганкелизации матрицы M назовём

$$\text{AHE} = \|M - \text{Hankel}(M)\|_2$$

Определение

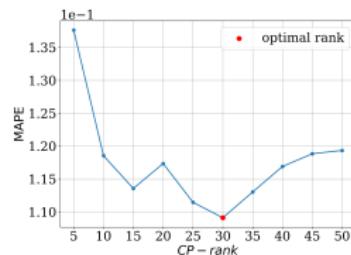
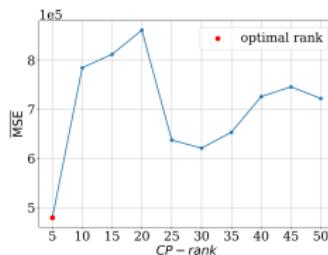
Относительной ошибкой ганкелизации матрицы M назовём

$$\text{RHE} = \frac{\text{AHE}}{\|M\|_2}$$

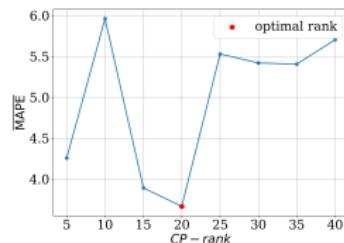
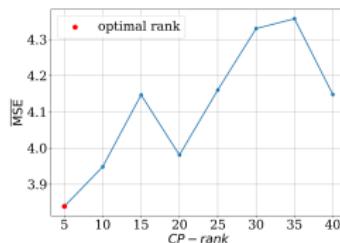
AHE имеет смысл суммарного стандартного отклонения антидиагоналей матрицы. RHE её более интерпретируемая модификация.

Данные метрики будут применяться к матрицам группировки факторов C_i .

Зависимость качества прогноза от CP-ранга



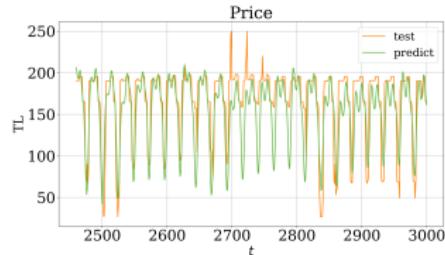
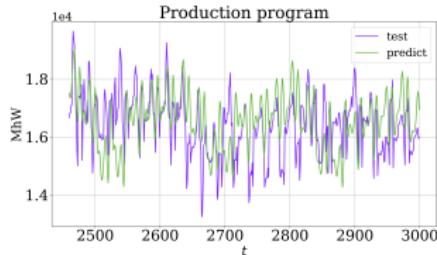
Данные электроэнергии.



Данные акселерометрии.

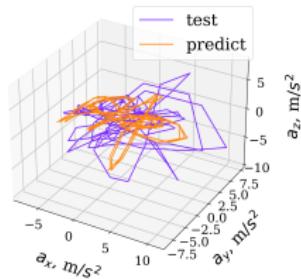
Отдельно выделен оптимальный ранг. Наблюдается эффект переобучения при возрастании ранга.

Визуализация прогноза методом tSSA

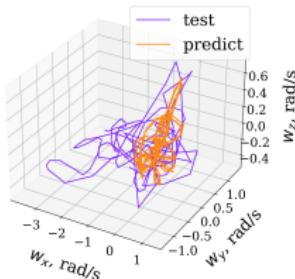


Данные электроэнергии. CPD-ранг = 30.

Accelerometer



Gyroscope



Данные акселерометрии. CPD-ранг = 20.

Тензорный ранг влияет на порядок модели прогноза.

Качество прогноза моделей

Данные электроэнергии.

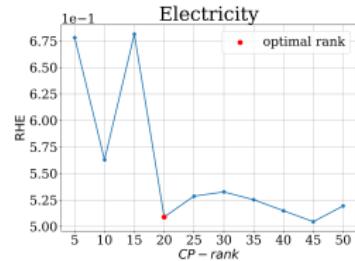
	<i>tSSA</i>	<i>mSSA</i>	<i>VAR</i>	<i>RNN</i>
MSE _{Production} , 10 ⁶	1.24	1.51	7.81	2.70
MSE _{Price} , 10 ³	0.88	1.03	4.85	30.0
MSE, 10 ⁶	0.62	0.75	3.91	135.00
MAPE _{Production}	0.054	0.060	0.137	0.999
MAPE _{Price}	0.164	0.170	0.360	1.004
MAPE	0.109	0.115	0.249	1.002

Данные акселерометрии.

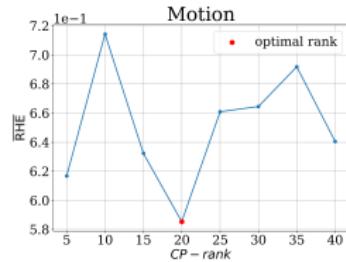
	<i>tSSA</i>	<i>mSSA</i>	<i>VAR</i>	<i>RNN</i>
MSE _{Accel}	7.351	6.980	8.108	6.604
MSE _{Gyro}	0.610	0.636	0.631	0.639
MSE	3.981	3.808	4.370	3.622
MAPE _{Accel}	3.558	3.516	3.370	1.747
MAPE _{Gyro}	3.773	3.943	10.427	5.641
MAPE	3.666	3.730	6.899	3.694

Рассматриваемый метод почти везде показал наилучшие результаты, *mSSA* был близок по точности. Метод *VAR* оказался нестабилен на выбранном горизонте прогнозирования, для *RNN* необходима большая длина предыстории.

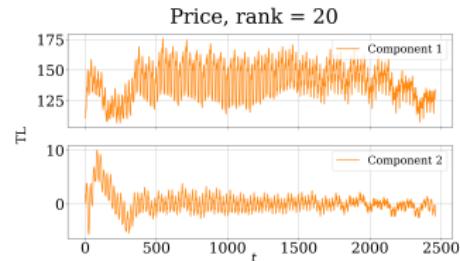
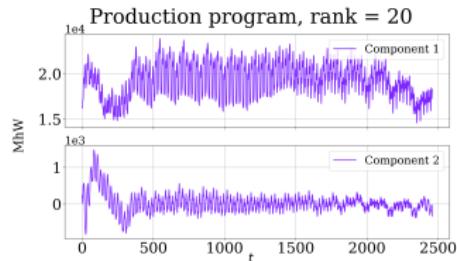
Анализ качества декомпозиции



(а) Данные электроэнергии. (b) Данные акселерометрии.



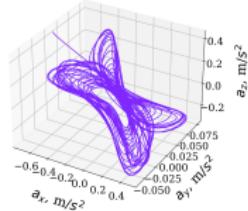
Отдельно выделен оптимальный ранг. Аналогичное падение качества при росте ранга. Разложение на две компоненты:



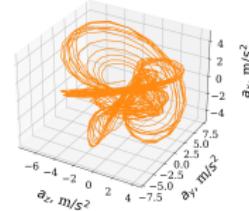
Данные электроэнергии.

Визуализация компонент разложения tSSA

Component 1

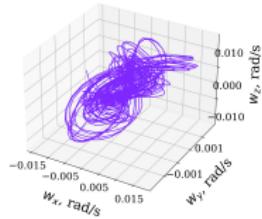


Component 2

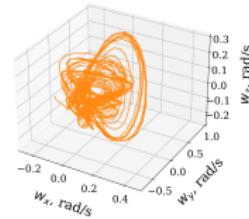


Данные акселерометра.

Component 1



Component 2



Данные гироскопа.

Компоненты наглядно показывают эффект общего базиса сигналов, а также нетривиальную динамику порождения данных.

Качество декомпозиции моделей

Данные электроэнергии.

	tSSA	mSSA
RHE _{Production}	0.507	0.308
RHE _{Price}	0.511	0.31
RHE	0.509	0.309

Данные акселерометрии.

	tSSA	mSSA
RHE _{Accel}	0.438	0.394
RHE _{Gyro}	0.732	0.468
RHE	0.585	0.431

Рассматриваемый метод немного уступает mSSA в качестве декомпозиции. Сложность задачи аддитивного разложения для tSSA является основным направлением развития метода.

- ➊ Разработан тензорный метод построения прогноза и декомпозиции многомерных временных рядов
- ➋ Доказана теорема о NP-сложности оптимальной декомпозиции, выведена формула и модель предсказания
- ➌ Предложенный подход успешно протестирован на реальных данных и сопоставлен с другими моделями
- ➍ Работа подготовлена к подаче в журнал *Multimedia Tools and Applications* под именем «*Tensor decomposition and forecast for multidimensional time series*»