Восстановление траектории движения руки по видео

Владимиров Эдуард
vladimirov.ea@phystech.edu

Курдюкова Антонина
kurdiukova.ad@phystech.edu

Исаченко Роман
roman.isachenko@phystech.edu

Стрижов Вадим
strijov@phystech.edu

17 мая 2022 г.

Аннотация

В работе решается задача прогнозирования временного ряда со сложной структурой. Под сложной структурой понимается наличие нелинейных зависимостей и варьирующийся период. Требуется найти причинно-следственные связи между временными рядами. Для этого предлагается снизить размерность траекторных пространств. В работе предложен новый способ согласованного снижения размерности многомерных временных рядов. Он объединяет метод частичных наименьших квадратов и метод перекрестных отображений Сугихары. Для демонстрации результатов работы решается задача восстановления траектории движения руки по видео.

Ключевые слова: временной ряд · фазовая траектория · траекторное подпространство · сходящееся перекрёстное отображение · частичные наименьшие квадраты

1 Введение

В данной работе решается задача прогнозирования временного ряда на основе других временных рядов. Одна из трудностей задачи заключается в обнаружении связи между временными рядами и исключении несвязанных временных рядов из прогностической модели. Решение этой проблемы повышает её качество.

В данной работе применяется метод сходящегося перекрёстного отображения (convergent cross mapping, CCM) или метод Сугихары [1, 2], который эффективен для временных рядов, порождённых динамической системой. Он основан на сравнении ближайших соседей в траекторном пространстве временного ряда $\mathbf{x}$, полученных с помощью временного ряда $\mathbf{y}$.

При построении прогностической модели используется траекторная матрица (или матрица сдвига), описывающая фазовое пространство временного ряда. Например, в методе анализа спектральных компонент (singular spectrum analysis, SSA) [3, 4, 5] прогноз временного ряда основан на спектральном разложении ковариационной матрицы, полученной по траекторной матрице. В CCM матрицы сдвига используются для проверки наличия липшицева отображения между траекторными пространствами.

Однако размерность траекторного пространства может оказаться чрезмерно высокой, что приводит к неустойчивости прогностической модели. В таком случае необходимо снизить размерность траекторного пространства путём построения проекции фазовой траектории в некоторое подпространство. Для CCM нет конкретного способа выбрать подпространство, в котором аппроксимируется фазовая траектория. В работе [6] эта проблема решается с помощью сферической регрессии. Согласно этому методу, информация об искомом подпространстве извлекается из множества эмпирических направлений $\{\mathbf{x}_i - \mathbf{x}_j \mid i < j\}$, где $\mathbf{x}_i$ — элементы траекторного пространства в сферических координатах. В работе [7] используется автоматический выбор пары главных компонент. Идея заключается в сравнении спектральных плотностей главных компонент. Также используется простой перебор по главным компонентам [8].

Метод проекции в латентное пространство (partial least squares, PLS) [9, 10] отбирает наиболее значимые признаки и строит новые как их линейные комбинации. Это позволяет получить простую, точную и устойчивую прогностическую модель. Наряду с PLS используется метод канонического анализа корреляции (canonical correlation analysis, CCA) [11]. Он похож на PLS за исключением того, что первый метод максимизирует ковариацию между проекциями, а последний — корреляцию. Недостатком этих моделей является их низкая точность при оценивании нелинейных зависимостей между данными. Разработаны нелинейные модели PLS [12] и CCA [13]. В данной статье используется модель PLS-Autoencoder [14], которая преобразует исходные данные с помощью автоэнкодеров.

В теоретической части работы показано, как можно применить метод Сугихары для снижения размерности траекторного пространства и как объединить идеи методов PLS и CCM. Для осуществления последней цели введена новая метрика согласованности латентных проекций.

В качестве модели для предсказания временного ряда по набору временных рядов используется алгоритм локально взвешенного глобального линейного отображения (sequential locally weighted global linear map, SMap) [15].

Эксперимент проводится на наборе собранных вручную данных. Он представляет собой совокупность ключевых точек, полученных по видео движения человека, а также показания акселерометра и гироскопа, снятые с руки человека. В эксперименте строится прогноз временных рядов, использующий обнаруженные связанные компоненты временных рядов.

2 Постановка задачи

Пусть значения исходного многомерного временного ряда

$$\mathbf{S}_y(t) = [S_y^1(t), \dots, S_y^r(t)]^\mathsf{T}$$

доступны в моменты времени $t = 1, 2, \dots, n$. Предполагается, что на значения $\mathbf{S}_y(t)$ оказывает влияние набор вспомогательных временных рядов $S_x^1(t), \dots, S_x^m(t)$.

Необходимо предсказать значения исходного временного ряда $\mathbf{S}_y(t)$ в моменты времени $n+1, \dots, n+p$. Предполагается, что значения вспомогательных временных рядов доступны в тот период времени, на который осуществляется предсказание временного ряда $\mathbf{S}_y(t)$.

Для вычисления будущих значений временного ряда требуется определить функциональную зависимость, отражающую связь между прошлыми значениями $\mathbf{S}_y(t)$ и будущими, а также принимающую во внимание влияние вспомогательных временных рядов $S_x^1(t), \dots, S_x^m(t)$.

Опр. 2.1. Моделью прогнозирования с учётом внешних факторов называется функция:

$$\mathbf{S}_y(t) = \mathbf{F}(\mathbf{w}, \mathbf{S}_y(t-1), \dots, S_x^1(t), \dots, S_x^m(t), \dots) + \boldsymbol{\varepsilon}_t.$$

Требуется создать такую модель, для которой среднее квадратичное отклонение истинного значения от прогнозируемого стремится к минимальному для заданного p.

$$\widehat{E} = \frac{1}{p} \sum_{i=n+1}^{n+p} ||\boldsymbol{\varepsilon}_i||_2^2 \to \min_{\mathbf{w}}.$$

Особенность данной задачи заключается в том, что размер m набора временных рядов довольно велик и что среди рядов $S_y^1(t), \dots, S_y^m(t)$ есть много сильно скореллированных. Поэтому использование всего набора для прогнозирования временного ряда $S_x(t)$ приводит к низкому качеству прогноза.

Один из способов решения этой проблемы заключается в выборе фиксированного числа временных рядов, оказывающих наибольшее влияние на целевую переменную, с помощью метода CCM. Для пары временных рядов

$$(S_x^i(t), S_y^j(t)) \quad i = 1, \dots, m \quad j = 1, \dots r$$

он определяет меру воздействия временного ряда $S_x^i(t)$ на целевую переменную $S_y^j(t)$. Далее выбираем временные ряды из набора с максимальной мерой воздействия.

2.1 Метод CCM

Определим траекторную матрицу временного ряда $\mathbf{x} = [x_1, \dots, x_n]$ следующим образом:

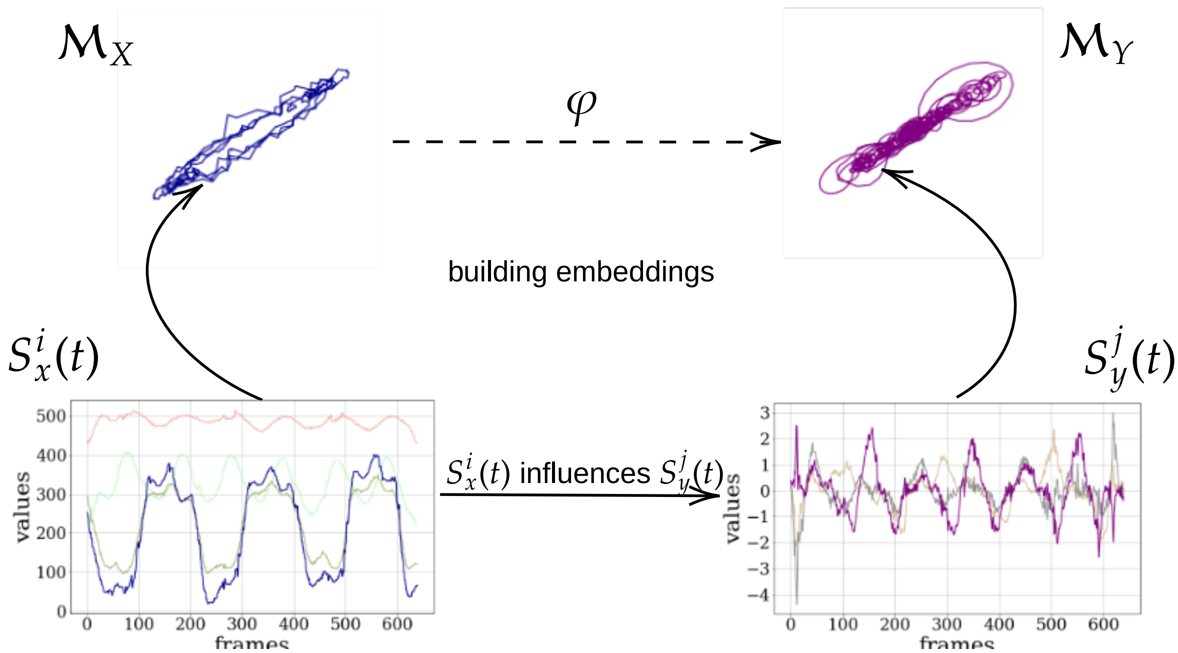


Рис. 1: Применение метода CCM для выбора наиболее значимых компонент временного ряда

$$\mathbf{H}_\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_\tau \\ x_2 & x_3 & \dots & x_{\tau+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N+1} & \dots & x_n \end{bmatrix},$$

где N — число задержек, $\tau = n - N + 1$.

Обозначим i-ый столбец матрицы $\mathbf{H}_\mathbf{x}$ за $\mathbf{x}_i$. Матрица $\mathbf{H}_\mathbf{x}$ принимает вид:

$$\mathbf{H}_\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_\tau], \qquad \mathbf{x}_i = [x_i, x_{i+1}, \dots, x_{i+N-1}]^\mathsf{T}$$

Заметим, что все векторы $\mathbf{x}_t$ принадлежат N-мерному траекторному пространству $\mathbb{H}_\mathbf{x} \subseteq \mathbb{R}^N$ ряда $\mathbf{x}$ и образуют фазовую траекторию $\mathbf{x}(t) \in \mathbb{R}^N$.

Для обнаружения зависимости между временными рядами $\mathbf{x}$ и $\mathbf{y}$ возьмём элемент $\mathbf{x}_0$ из траекторного пространства $\mathbb{H}_\mathbf{x}$ и найдём k ближайших соседей в этом же пространстве. Обозначим их временные индексы (от ближнего к дальнему) $t_1, \dots, t_k$.

Так как оба временных ряда определены на одной временной оси, то по значению временного ряда $\mathbf{x}$ в момент времени $t_0 \in \{1, \dots, n\}$ можно однозначно получить значение временного ряда $\mathbf{y}$ в тот же момент времени, и наоборот. Введём отображение из $\mathbb{H}_\mathbf{x}$ в

$\mathbb{H}_{\mathbf{y}}$ следующим образом:

$$\varphi : \mathbf{x}_0 \mapsto \widehat{\mathbf{y}_0} = \sum_{i=1}^{k} w_i \mathbf{y}_{t_i}, \qquad w_i = \frac{u_i}{\sum\limits_{j=1}^{k} u_j}, \qquad u_i = \exp(-||\mathbf{x}_0 - \mathbf{x}_{t_i}||).$$

Опр. 2.2. Временные ряды $\mathbf{x}$ и $\mathbf{y}$ называются **связанными**, если отображение φ является липшицевым:

$$\rho_{\mathbb{H}_{\mathbf{y}}}(\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j)) \leqslant C \rho_{\mathbb{H}_{\mathbf{x}}}(\mathbf{x}_i, \mathbf{x}_j) \qquad \mathbf{x}_i, \mathbf{x}_j \in \mathbb{H}_{\mathbf{x}}.$$

Для проверки наличия связанности введём метрическую функцию близости векторов в окрестностях $U_k(\mathbf{x}_{t_0})$ и $U_k(\mathbf{y}_{t_0})$:

$$L(\mathbf{x}, \mathbf{y}) = \frac{R(U_k(\mathbf{x}_{t_0}))}{R(U_k(\mathbf{y}_{t_0}))}, \qquad R(U_k(\mathbf{x}_{t_0})) = \frac{1}{k} \sum_{i=1}^{k} \rho_{\mathbb{H}_{\mathbf{x}}}(\mathbf{x}_{t_0}, \mathbf{x}_{t_j}). \tag{1}$$

Если $L(x, y)$ больше заданного порога $C(n)$, то временной ряд $\mathbf{y}$ зависит от временного ряда $\mathbf{x}$.

2.2 Метод PLS

Другой способ решения поставленной ранее проблемы заключается в согласованном снижении размерности Метод частичных наименьших квадратов восстанавливает связь между наборами данных $\mathbf{X}$ и $\mathbf{Y}$. Матрицы объектов $\mathbf{X}$ и целевая матрица $\mathbf{Y}$ проецируются на латентное пространство $\mathbb{R}^l$ меньшей размерности следующим образом:

$$\underset{n\times d}{\mathbf{X}} = \underset{n\times K}{\mathbf{T}} \cdot \underset{K\times d}{\mathbf{P}^{\mathsf{T}}} + \underset{n\times d}{\mathbf{E}}$$

$$\underset{n\times s}{\mathbf{Y}} = \underset{n\times K}{\mathbf{U}} \cdot \underset{K\times s}{\mathbf{Q}^{\mathsf{T}}} + \underset{n\times s}{\mathbf{F}},$$

где $\mathbf{T}$ и $\mathbf{U}$ — матрицы описания объектов и исходов в латентном пространстве, $\mathbf{P}$ и $\mathbf{Q}$ — матрицы перехода из латентного пространства в исходное, $\mathbf{E}$ и $\mathbf{F}$ — матрицы остатков.

Псевдокод метода регрессии PLS приведен в алгоритме 1. Алгоритм итеративно на каждом из K шагов вычисляет по одному столбцу $\mathbf{t}_k, \mathbf{u}_k, \mathbf{p}_k, \mathbf{q}_k$ матриц $\mathbf{T}, \mathbf{U}, \mathbf{P}, \mathbf{Q}$ соответственно. После вычисления следующего набора векторов из матриц $\mathbf{X}, \mathbf{Y}$ вычитаются очередные одноранговые аппроксимации.

Функция преобразования исходных данных имеет вид:

$$f(\mathbf{X}) = \mathbf{X}\mathbf{W}_{\mathbf{x}} \qquad g(\mathbf{Y}) = \mathbf{Y}\mathbf{W}_{\mathbf{y}},$$

где матрицы весов $\mathbf{W}_{\mathbf{x}} \in \mathbb{R}^{d\times K}$ и $\mathbf{W}_{\mathbf{y}} \in \mathbb{R}^{s\times K}$ находятся путём максимизации выборочной ковариации:

$$(\mathbf{W}_{\mathbf{x}}, \mathbf{W}_{\mathbf{y}}) = \underset{\mathbf{W}_{\mathbf{x}}, \mathbf{W}_{\mathbf{y}}}{\operatorname{argmax}} \operatorname{Cov}(\mathbf{X}\mathbf{W}_{\mathbf{x}}, \mathbf{Y}\mathbf{W}_{\mathbf{y}})$$

Из алгоритма PLS можно получить явный вид матриц $\mathbf{W_x}$ и $\mathbf{W_y}$. Заметим, что:

$$\mathbf{X} \cdot \mathbf{A}(\mathbf{P}^\mathsf{T}\mathbf{A})^{-1} = (\mathbf{T}\mathbf{P}^\mathsf{T}\mathbf{A} + \mathbf{E}\mathbf{A})(\mathbf{P}^\mathsf{T}\mathbf{A})^{-1} \approx \mathbf{T},$$

где матрица $\mathbf{A}$ образована из столбцов $\mathbf{a}_k$. Аналогично, $\mathbf{Y} \cdot \mathbf{B}(\mathbf{Q}^\mathsf{T}\mathbf{B})^{-1} \approx \mathbf{U}$, где матрица $\mathbf{B}$ образована из столбцов $\mathbf{b}_k$. Таким образом:

$$\mathbf{W_x} = \mathbf{A}(\mathbf{P}^\mathsf{T}\mathbf{A})^{-1}, \qquad \mathbf{W_y} = \mathbf{B}(\mathbf{Q}^\mathsf{T}\mathbf{B})^{-1}.$$

Algorithm 1 Canonical PLS

Require: $\mathbf{X} \in \mathbb{R}^{n\times d}$, $\mathbf{Y} \in \mathbb{R}^{n\times s}$, $K \in \mathbb{N}$

Ensure: $\mathbf{T}$, $\mathbf{U}$, $\mathbf{P}$, $\mathbf{Q}$

Нормировать матрицы $\mathbf{X}$ и $\mathbf{Y}$ по столбцам
$\mathbf{X}_1 \leftarrow \mathbf{X}$
$\mathbf{Y}_1 \leftarrow \mathbf{Y}$
for $k = 1, \dots, K$ **do**
  if $\mathbf{X}_k^T\mathbf{Y}_k = 0$ **then**
    break
  end if
  Вычислить $\mathbf{a}_k \in \mathbb{R}^d$ и $\mathbf{b}_k \in \mathbb{R}^s$, первые
  левые и правые сингулярные вектора матрицы $\mathbf{X}_k^T\mathbf{Y}_k$.
  Из определения следует, что $(\mathbf{a}_k, \mathbf{b}_k) = \underset{\mathbf{a},\mathbf{b}}{\operatorname{argmax}} \operatorname{Cov}(\mathbf{X}_k\mathbf{a}, \mathbf{Y}_k\mathbf{b})$.
  $\mathbf{t}_k \leftarrow \mathbf{X}_k\mathbf{a}_k$
  $\mathbf{u}_k \leftarrow \mathbf{Y}_k\mathbf{b}_k$
  $\mathbf{p}_k^T = (\mathbf{t}_k^T\mathbf{t})^{-1}\mathbf{t}_k^T\mathbf{X}_k$ ▷ Вектора $\mathbf{p}_k$ образуют матрицу $\mathbf{P}$
  $\mathbf{q}_k^T = (\mathbf{u}_k^T\mathbf{u}_k)^{-1}\mathbf{u}_k^T\mathbf{Y}_k$ ▷ Вектора $\mathbf{q}_k$ образуют матрицу $\mathbf{Q}$
  $\widehat{\mathbf{X}_k}(\mathbf{t}_k) = \mathbf{t}_k\mathbf{p}_k^T$ ▷ Регрессируем $\mathbf{X}_k$ по $\mathbf{t}_k$
  $\widehat{\mathbf{Y}_k}(\mathbf{u}_k) = \mathbf{u}_k\mathbf{q}_k^T$ ▷ Регрессируем $\mathbf{Y}_k$ по $\mathbf{u}_k$
  $\mathbf{X}_{k+1} \leftarrow \mathbf{X}_k - \widehat{\mathbf{X}_k}(\mathbf{t}_k)$
  $\mathbf{Y}_{k+1} \leftarrow \mathbf{Y}_k - \widehat{\mathbf{Y}_k}(\mathbf{u}_k)$
end for

2.3 Методы PLS-Autoencoder и PLS-CCM

Основной минус классического метода PLS заключается в низком качестве при работе с данными, которые имеют сложные нелинейные зависимости. По этой причине были разработаны расширения линейного метода PLS, которые преобразуют входные данные с помощью нейронных сетей. Одним из таких расширений является метод PLS-Autoencoder. В качестве параметрической функций, которые переводят исходные данные в латентное пространство и обратно, выступают нейронные сети. Эти функции называются энкодером и декодером соответственно.

$$
\begin{tikzcd}
\underset{n\times d}{\mathbf{X}} \arrow[d, bend right, "\varphi_{enc}"'] & & \underset{n\times s}{\mathbf{Y}} \arrow[d, bend left, "\psi_{enc}"] \\
\underset{n\times K}{\mathbf{T}} \arrow[u, "\varphi_{dec}"'] \arrow[rr, leftrightarrow, "\mathcal{L}_{cons}"] & & \underset{n\times}{\mathbf{U}} \arrow[u, "\psi_{dec}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\underset{n\times d}{\mathbf{X}} \arrow[d, bend right, "\varphi_{enc}"'] \arrow[rr, leftrightarrow, "CCM_{XY}"] & & \underset{n\times s}{\mathbf{Y}} \arrow[d, bend left, "\psi_{enc}"] \\
\underset{n\times K}{\mathbf{T}} \arrow[u, "\varphi_{dec}"'] \arrow[rr, leftrightarrow, "\mathcal{L}_{fc} + CCM_{TU}"] & & \underset{n\times K}{\mathbf{U}} \arrow[u, "\psi_{dec}"]
\end{tikzcd}
$$

3 Вычислительный эксперимент

Целью эксперимента является сравнение различных стратегий снижения размерности целевого пространства. Важной его частью является изучение результатов алгоритма прогнозирования временного ряда, применённого к элементам пространства фазовых траекторий и траекторного подпространства меньшей размерности.

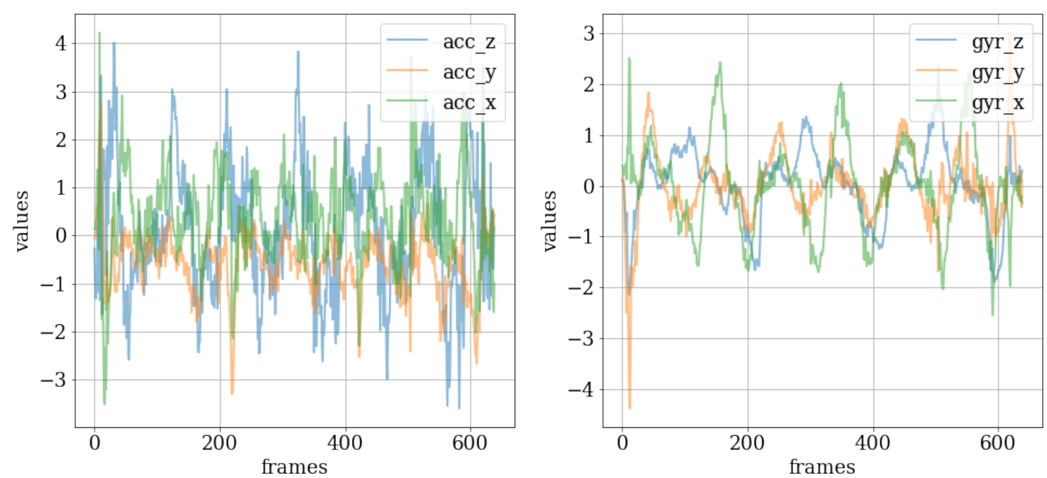
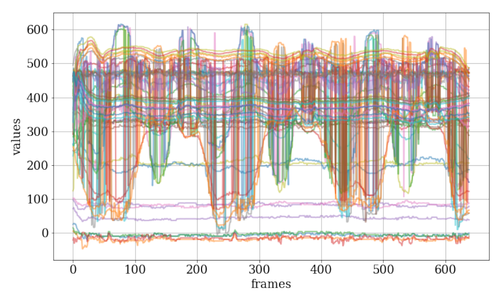


Рис. 2: Данные акселерометра и гироскопа

Первоначальные данные представляют собой набор видеороликов, на которых выполняются различные движения руками (циклические и хаотические), а также показания акселерометра и гироскопа частотой в 100 Герц, закреплённых на одной из рук. Далее по видеоряду с помощью фреймворка alphapose [16, 17, 18] получаются координаты конечностей, а именно 68 ключевых точек. Затем из полученного временного ряда исключаются сильно скореллированные компоненты. После этого полученные многомерные временные ряды приводятся к одной временной шкале с помощью удаления элементов более длинного временного ряда.

Перед началом эксперимента зафиксируем следующие переменные: N — размерность траекторного пространства, k — число ближайших соседей, рассматриваемых в CCM, n_{tr} — размер обучающей выборки, E_{vid} — число признаков, полученных из видеоряда, которые будут использованы в алгоритме. Для каждой пары компонент, взятых из разных временных рядов, применим метод CCM. В результате получим матрицу корреляций, в которой на i-ой строке и j-ом столбце стоит коэффициент корреляции Пирсона между i-ой компонентной "приборного"временного ряда и j-ой компонентой ряда, восстановленного по видео. После этого для каждого целевого признака выбираем E_{vid} видео-признаков, обладающих максимальной корреляцией в соответствующей строке. Затем на основе этих признаков обучается алгоритм.

(a) Результат работы фреймворка alphapose



(b) Данные кейпоинтов, полученные по видео

4 Анализ ошибки

Таблица 1: Сравнение ошибки (MSE) предсказательной модели, применённой в траекторном пространстве и в его подпространстве, полученном CCM

	acc_z	acc_y	acc_x	gyr_z	gyr_y	gyr_x
space	1.053 ± 2.223	0.401 ± 0.833	0.483 ± 0.825	0.084 ± 0.537	0.090 ± 0.094	0.063 ± 0.295
subspace	0.315 ± 0.461	0.043 ± 0.051	0.150 ± 0.177	0.001 ± 0.001	0.015 ± 0.031	0.001 ± 0.003

Для начала сравним качество предсказаний прогностической модели, применённой в траекторном пространстве и его подпространстве. В таблице 1 представлены среднеквадратичная ошибки предсказаний значений акселерометра и гироскопа по каждой из осей и их стандартные отклонения. На ней видно, что прогностическая модель, применённая в траекторном подпространстве, даёт более точные предсказания, поскольку большинство признаков исходного признакового пространства неинформативно.

Далее, рассмотрим различные методы снижения размерности траекторного пространства. В таблице 2 видно, что добавление данных из видеоряда повышает качество модели предсказания.

Таблица 2: Среднеквадратичное отклонение между истинными показаниями устройств и предсказаниями, полученными с помощью одного из методов снижения размерности

	Метод / Целевой признак	CCM	PLS	CCA
cyclic	acc_z	0.163	**0.040**	0.116
	acc_y	0.009	**0.007**	0.011
	acc_x	**0.044**	0.045	0.089
	gyr_z	**0.000**	0.001	0.001
	gyr_y	**0.002**	0.004	0.005
	gyr_x	0.009	**0.004**	**0.004**
chaotic	acc_z	**0.315**	0.416	0.416
	acc_y	**0.043**	0.045	0.429
	acc_x	**0.150**	0.177	0.221
	gyr_z	**0.001**	0.002	0.003
	gyr_y	**0.015**	0.022	0.061
	gyr_x	**0.001**	0.013	0.015

5 Заключение

В работе предложен метод обобщения методов PLS и CCA с помощью метода Сугихары путём построения эмбеддингов и выбора метрики для оценки качества аппроксимации. Проведён вычислительный эксперимент на данных устройств и видеоряда. Получено, что использование данных из видео повышает качество прогнозирования. Показано, что прогностическая модель менее устойчива в случае, когда та применяется в траекторном пространстве.

В дальнейшем планируется применить метод не к двумерным данным, которые соответствуют регулярным измерениям некоторой величины, а уже к спорадическим временным рядам. Это означает, что входными данными будут служить тензоры (???).

Список литературы

[1] G. Sugihara, B. Grenfell, and R. M. May. Distinguishing error from chaos in ecological time series. *Phil. Trans. Roy. Soc. London B*, 330(1257):235–51, 1990.

[2] George Sugihara and Robert M May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734–741, 1990.

[3] Nina Golyandina and D Stepanov. Ssa-based approaches to analysis and forecast of multidimensional time series. In *proceedings of the 5th St. Petersburg workshop on*

simulation, volume 293, page 298. St. Petersburg State University St. Petersburg, Russia, 2005.

[4] Nina Golyandina, Vladimir Nekrutkin, and Anatoly A Zhigljavsky. *Analysis of time series structure: SSA and related techniques*. CRC press, 2001.

[5] Anatoly Zhigljavsky. Singular spectrum analysis for time series: Introduction to this special issue. *Statistics and its Interface*, 3(3):255–258, 2010.

[6] Карина Равилевна Усманова, ЮИ Журавлёв, КВ Рудаков, and ВВ Стрижов. Аппроксимация фазовой траектории квазипериодических сигналов методом сферической регрессии. *Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика*, (4):40–46, 2020.

[7] Th Alexandrov and N Golyandina. Automatic extraction and forecast of time series cyclic components within the framework of ssa. In *Proceedings of the 5th St. Petersburg Workshop on Simulation*, pages 45–50. St. Petersburg State University St. Petersburg, 2005.

[8] Карина Равилевна Усманова and Вадим Викторович Стрижов. Модели обнаружения зависимостей во временных рядах в задачах построения прогностических моделей. *Системы и средства информатики*, 29(2):12–30, 2019.

[9] Roman Rosipal. Nonlinear partial least squares an overview. *Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques*, pages 169–189, 2011.

[10] Roman Rosipal and Nicole Kramer. Overview and recent advances in partial least squares. In *International Statistical and Optimization Perspectives Workshop Subspace, Latent Structure and Feature Selection*, pages 34–51. Springer, 2005.

[11] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12): 2639–2664, 2004.

[12] S Joe Qin and Thomas J McAvoy. Nonlinear pls modeling using neural networks. *Computers & Chemical Engineering*, 16(4):379–391, 1992.

[13] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International conference on machine learning*, pages 1247–1255. PMLR, 2013.

[14] Nicholas Polson, Vadim Sokolov, and Jianeng Xu. Deep learning partial least squares. *arXiv preprint arXiv:2106.14085*, 2021.

[15] George Sugihara. Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 348(1688):477–495, 1994.

[16] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. RMPE: Regional multi-person pose estimation. In *ICCV*, 2017.

[17] Jiefeng Li, Can Wang, Hao Zhu, Yihuan Mao, Hao-Shu Fang, and Cewu Lu. Crowdpose: Efficient crowded scenes pose estimation and a new benchmark. *arXiv preprint arXiv:1812.00324*, 2018.

[18] Yuliang Xiu, Jiefeng Li, Haoyu Wang, Yinghong Fang, and Cewu Lu. Pose Flow: Efficient online pose tracking. In *BMVC*, 2018.