

«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (национальный
исследовательский университет)»

Физтех-школа прикладной математики и информатики

Кафедра «Интеллектуальные системы»

Курдюкова Антонина Дмитриевна

Снижение размерности фазового пространства в задачах канонического корреляционного анализа

03.03.01 – Прикладные математика и физика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Научный руководитель:

д.ф.-м.н. Стрижов Вадим Викторович

Москва

2022

Содержание

| | | |
|----------|---|-----------|
| 1 | Введение | 5 |
| 1.1 | Введение | 5 |
| 2 | Теоретическая часть | 7 |
| 2.1 | Метод сходящихся перекрестных отображений | 7 |
| 2.2 | Метод проекций на латентные структуры | 8 |
| 2.3 | Вычислительный эксперимент | 10 |
| 2.4 | Постановка задачи | 10 |
| 3 | Заключение | 11 |

Аннотация

Данная работа посвящена задаче снижения размерности фазового пространства методами канонического корреляционного анализа. Исследуется связь между методом канонического корреляционного анализа и методом сходящихся перекрестных отображений Сугихары. Вид прогностических моделей представим в виде условия принадлежности двух аттракторов, восстанавливаемых в исходном и целевом фазовых пространствах, к общей динамической системе. В работе рассмотрены методы PLS-CCA, нелинейный метод CCA, seq2seq, Neural ODE. Сформулирован вариант теоремы о вложениях Такенса для проверки того, что метод канонического корреляционного анализа или другой метод прогноза удовлетворяет условиям Сугихары. Решается прикладная задача в теоретической постановке. Рассматривается видеоряд ходьбы человека с акселерометром на руке.

Ключевые слова: *снижение размерности, фазовое пространство, аттрактор, CCM, теорема Такенса о вложениях*

1 Введение

1.1 Введение

Решается задача прогнозирования сигналов походки человека. Такие сигналы обладают сложной структурой, то есть имеются зависимости и изменяющийся период. Рассматриваются два связанных пространства. Например, акселерометр и гироскоп одного мобильного устройства, два акселерометра в правой и левой руке человека, траектория движения руки, восстановленная по видеоряду движения человека, и акселерометр на этой руке.

Для улучшения качества прогноза, а также для упрощения прогностической модели предлагается учесть зависимость между временными рядами, а также перейти в пространство меньшей размерности. Снижение размерности позволит учитывать внутреннее низкоразмерное представление временных рядов в прогностической модели.

Для определения наличия связи между временными рядами используется метод сходящегося перекрестного отображения (convergent cross mapping, CCM) [1, 2]. Метод CCM проверяет, насколько близки точки фазового пространства временного ряда s_1 , соответствующие ближайшим соседям ряда s_2 . Иными словами, существует ли взаимно однозначное соответствие, которое отображает окрестность фазовой траектории s_1 в окрестность фазовой траектории s_2 .

Для снижения размерности траекторного пространства используются метод проекций на латентные структуры (partial least squares PLS) [3, 4], нелинейный PLS [5], seq2seq, NeuralODE. Снижение размерности позволяет сделать прогностическую модель более устойчивой, изучить связь между главными компонентами временных рядов, а также найти траекторное подпространство, в котором удастся обнару-

жить связь между временными рядами.

В работе исследуется связь между методами корреляционного анализа и методом сходящегося перекрестного отображения. Для ССМ нет способа выбора собственного подпространства, в котором аппроксимируется многообразие компакта и работает прогностическая модель. На текущий момент выбор собственного пространства осуществляется перебором по главным компонентам, например в [6]. Работа Исаченко [..] по PLS дает возможность перенести методы выбора подпространства с PLS на ССМ.

Требуется построить прогностическую модель, связывающую метод сходящегося перекрестного отображения и методы канонического корреляционного анализа. Целью работы является построение более простой модели, работающей не хуже уже существующих моделей прогнозирования временных рядов.

Определение 1 *Динамическая система – множество элементов, для которого задана функциональная зависимость между временем и положением в фазовом пространстве каждого элемента системы*

Динамическая система представляет собой такую математическую модель некоего объекта, процесса или явления, в которой пренебрегают «флуктуациями и всеми другими статистическими явлениям».

Определение 2 *Многообразие – хаусдорфово топологическое пространство со счётной базой, каждая точка которого обладает окрестностью, гомеоморфной евклидову пространству \mathbb{R}^n*

2 Теоретическая часть

Пусть $s_1 = \{s_i^1\}_{i=1}^{N_1}$ и $s_2 = \{s_i^2\}_{i=1}^{N_2}$ — заданные временные ряды. Опишем, как строится фазовое пространства временного ряда. Строится ганкелева матрица для ряда s :

$$\mathbf{H}_1 = \begin{bmatrix} s_1 & \dots & s_{n_1} \\ s_2 & \dots & s_{n_1+1} \\ \vdots & \ddots & \vdots \\ s_{k_1} & \dots & s_{N_1} \end{bmatrix}^T = \left[\mathbf{s}_1^1, \mathbf{s}_2^1, \dots, \mathbf{s}_k^1 \right], \quad k_1 = N_1 - n_1 + 1,$$

где n — ширина окна. Аналогично для временного ряда s_2 . Тогда вектора $\mathbf{s}_1^1, \mathbf{s}_2^1, \dots, \mathbf{s}_k^1$ образуют фазовую траекторию, или, иными словами, аттрактор \mathbf{M}_1 временного ряда s_1 . На эти же вектора натянуто фазовое пространство размерности n_1 временного ряда s_1 .

Определение 3 *Фазовое пространство динамической системы — совокупность всех допустимых состояний динамической системы.*

Определение 4 *Траектория динамической системы в фазовом пространстве — последовательность состояний*

Определение 5 *Аттрактор — компактное подмножество фазового пространства динамической системы, все траектории из некоторой окрестности которого стремятся к нему при времени, стремящемся к бесконечности.*

2.1 Метод сходящихся перекрестных отображений

Метод сходящихся перекрестных отображений (convergent cross mapping, CMM) используется для исследования временных рядов на наличие причинно-следственной

связи. Корреляция не подразумевает причинно-следственную связь между рядами. Метод основан на теореме Такенса о вложениях. В общем случае многообразие аттрактора динамической системы может быть восстановлено по одной наблюдаемой \mathbf{X} .

Согласно методу временной ряд s_1 может быть восстановлен по ряду s_2 только если временной ряд s_2 связан с рядом s_1 . Временные ряды считаются связанными, если окрестность фазовой траектории \mathbf{x} временного ряда s_1 взаимно однозначно отображается в окрестность фазовой траектории \mathbf{y} ряда s_2 . Иными словами,

Определение 6 *Аттракторы \mathbf{M}_1 и \mathbf{M}_2 наблюдаемых \mathbf{X} и \mathbf{Y} , если \mathbf{X} и \mathbf{Y} принадлежат одной динамической системе.*

2.2 Метод проекций на латентные структуры

Метод проекций на латентные структуры PLS [3, 4] используют для нахождения фундаментальных зависимостей между двумя матрицами \mathbf{X} и \mathbf{Y} . Отбираются наиболее значимые признаки. Новые признаки являются их линейными комбинациями. Осуществляется переход в фазовое пространство меньшей размерности. Метод PLS позволяет найти фазовое подпространство, в котором наблюдается связь между главными компонентами исходных временных рядов. Это позволяет исследовать наличие связи между временными рядами.

Пусть $\mathbf{X} \in \mathbb{R}^{m \times n}$ и $\mathbf{Y} \in \mathbb{R}^{m \times r}$ — матрицы двух фазовых пространств, построенных по временному ряду \mathbf{s}_1 и \mathbf{s}_2 соответственно. Требуется построить прогноз временного ряда \mathbf{s}_2 с учетом связи с временным рядом \mathbf{s}_1 . Предполагается линейная зависимость между строками \mathbf{X} и \mathbf{Y} :

$$\mathbf{Y}_i = \mathbf{X}_i \cdot \boldsymbol{\Theta} + \boldsymbol{\epsilon} \quad \mathbf{Y}_i \in \mathbb{R}^r, \mathbf{Y}_i \in \mathbb{R}^n, i = 1, \dots, m, \quad (1)$$

где Θ — матрица весов линейной зависимости, ε — вектор ошибок.

Ошибка вычисляется по формуле:

$$S(\Theta, \mathbf{X}, \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X} \cdot \Theta\|_2^2 \quad (2)$$

Алгоритм PLS находит матрицы $\mathbf{T}, \mathbf{P}, \mathbf{Q}$, с помощью которых осуществляется переход в латентное пространство согласно формулам:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^\top + \mathbf{F} \quad (3)$$

$$\mathbf{Y} = \mathbf{T} \cdot \mathbf{Q}^\top + \mathbf{E} \quad (4)$$

Матрица \mathbf{T} наилучшим образом описывает \mathbf{X} и \mathbf{Y} . Ее столбцы ортогональны. Матрицами \mathbf{P} и \mathbf{Q} определяется переход из латентного пространства в исходное. Матрицы \mathbf{X} и \mathbf{Y} — матрицы невязок.

Алгоритм PLS также позволяет определить матрицу \mathbf{W} , с помощью которой рассчитывается матрица весов Θ :

$$\Theta = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1} \mathbf{Q}^\top \quad (5)$$

2.3 Вычислительный эксперимент

Вычислительный эксперимент проводился на данных [7].

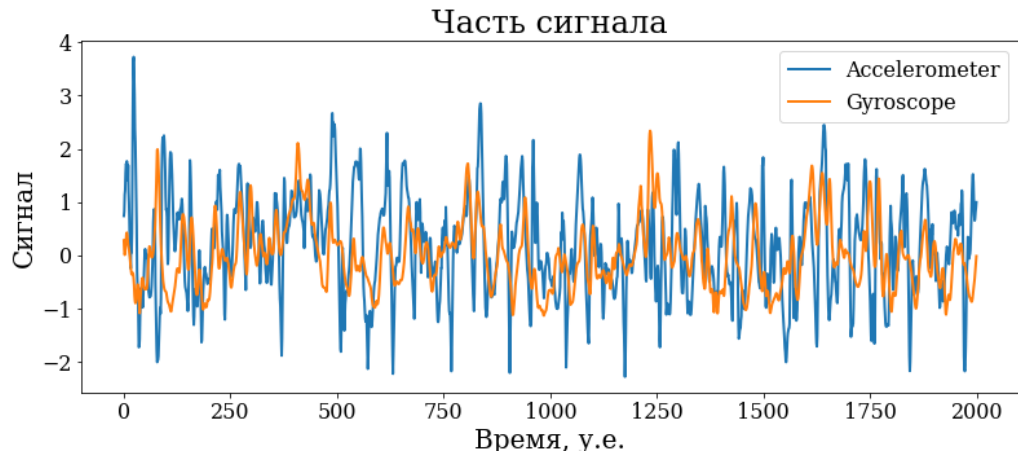


Рис. 1: Сигнал акселерометра и гироскопа походки человека

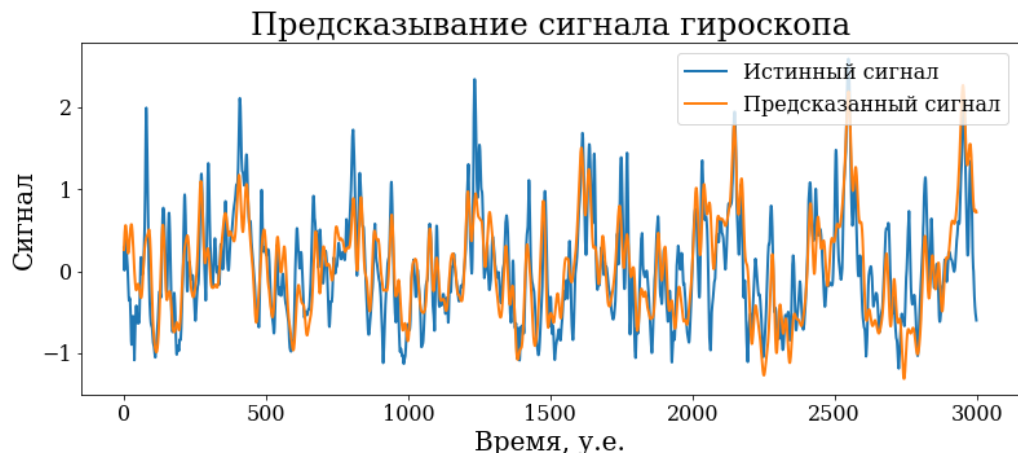


Рис. 2: Предсказание сигнала гироскопа по акселерометру, алгоритм PLS

2.4 Постановка задачи

3 Заключение

Список литературы

- [1] 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.
- [2] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.
- [3] Paul Geladi. Notes on the history and nature of partial least squares (pls) modelling. *Journal of Chemometrics*, 2(4):231–246, 1988.
- [4] Agnar Höskuldsson. Pls regression methods. *Journal of chemometrics*, 2(3):211–228, 1988.
- [5] Фарух Юрьевич Яушев, Роман Владимирович Исаченко, and Вадим Викторович Стрижов. Модели согласования скрытого пространства в задаче прогнозирования. *Системы и средства информатики*, 31(1):4–16, 2021.
- [6] 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.
- [7] Данные акселерометра и гироскопа: <https://github.com/Intelligent-Systems-Phystech/Kurdyukova-BS-Thesis/tree/master/code/data>.