

Модели согласования скрытого пространства в задаче корреляционного анализа

Роберт Сафиуллин

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

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Москва, 2023

Задача декодирования временного ряда

Задача

Дана пара синхронизированных рядов из двух фазовых пространств одной системы. Требуется построить устойчивую предсказательную модель для восстановления одного ряда по известным значениям второго

Проблема

Сложная структура рядов - наличие нелинейных зависимостей и высокая размерность данных.

Предлагается

Построить модели для получения низкоразмерной аппроксимации измерений состояния системы в исходном и целевом пространствах. Между полученными скрытыми представлениями построить согласующую модель для отображения одного пространства на другое.

Постановка задачи

Дано

$(\mathbf{X}, \mathbf{Y}) = \{x_i, y_i\}_{i=1}^N$ - выборка из сигналов.

$x_i, y_i \in \mathbb{R}^{a \times T}$

Строятся модели аппроксимации:

$$\phi(\mathbf{X}, \mathbf{w}) : \mathbb{R}^{a \times T} \rightarrow \mathbb{R}^h$$

$$\psi(\mathbf{Y}, \gamma) : \mathbb{R}^{a \times T} \rightarrow \mathbb{R}^p$$

для получения признаков описаний исходных рядов и модель согласования этих представлений:

$$\mathbf{f}(\phi(\mathbf{X}), \theta) : \mathbb{R}^h \rightarrow \mathbb{R}^p$$

Прогноз $\hat{\mathbf{Y}}$ ряда \mathbf{Y} :

$$\hat{\mathbf{Y}} = \mathbf{g}(\mathbf{f}(\phi(\mathbf{X})))$$

$$\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^{a \times T}$$

Функция ошибки:

$$\mathcal{L}(\mathbf{Y}, \mathbf{X}, \mathbf{w}, \theta) = \|\hat{\mathbf{Y}} - \mathbf{Y}\|_2^2$$

Симметрично положительно определенные (СПО) матрицы $\{\mathbf{C}_j\}_{j=1}^S \in \mathbb{R}^{a \times a}$ формируют риманово многообразие \mathbf{M} , расстояние между двумя точками $\mathbf{C}_1, \mathbf{C}_2$ которого определяется как:

$$\delta_R(\mathbf{C}_1, \mathbf{C}_2) = \|\log(\mathbf{C}_1^{-1}\mathbf{C}_2)\|_F = (\sum_{i=1}^n \log^2 \lambda_i)^{1/2}$$

где $\lambda_1, \dots, \lambda_n$ - положительные собственные значения матрицы $\mathbf{C}_1^{-1}\mathbf{C}_2$.

Для каждой ковариационной матрицы \mathbf{C}_i существует проекция \mathbf{S}_i в касательную плоскость в точке среднего геометрического \mathbf{C} :

$$\mathbf{S}_i = \mathbf{C}^{1/2} \log(\mathbf{C}^{-1/2} \mathbf{C}_i \mathbf{C}^{-1/2}) \mathbf{C}^{1/2} \quad (1)$$

$$\mathbf{C} = \arg \min_{\mathbf{C}} \sum_i \delta_R(\mathbf{C}, \mathbf{C}_i)^2 \quad (2)$$

Пусть $\mathbf{X} = \{x_i\}_{i=1}^N$ и $\mathbf{Y} = \{y_i\}_{i=1}^M$ обозначают временные ряды, полученные от двух различных пространств, согласованных таким образом, что они представляют собой отображения одной и той же системы. Пусть представления этих рядов в виде ковариационных матриц обозначены как $\{\mathbf{C}_i^{\mathbf{X}}\}_{i=1}^N$ и $\{\mathbf{C}_i^{\mathbf{Y}}\}_{i=1}^M$ соответственно.

Теорема 1. (Сафиуллин, 2023)

Эти представления лежат на одном и том же Римановом многообразии, но в разных его точках, определяющихся средними ковариационных матриц данных измерений.

Гипотеза

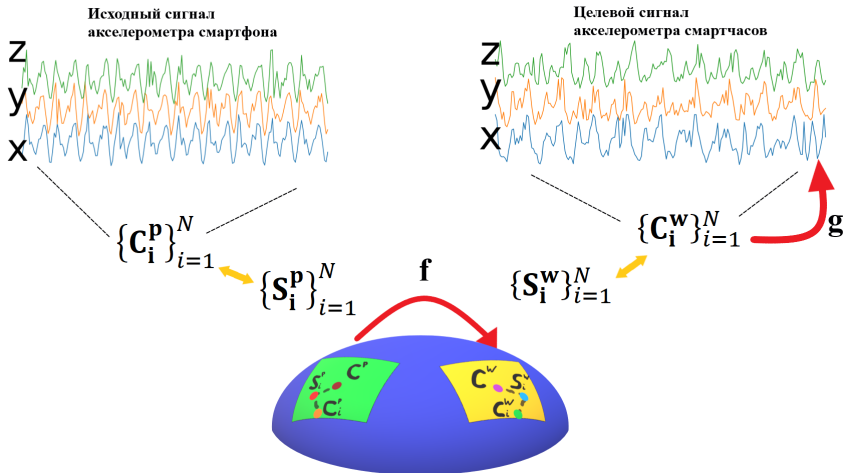
С помощью признакового описания:

$$\text{vec}(\mathbf{S}) = [\mathbf{S}_{1,1}; \sqrt{2}\mathbf{S}_{1,2}; \sqrt{2}\mathbf{S}_{1,3}; \mathbf{S}_{2,2}; \sqrt{2}\mathbf{S}_{2,3}; \mathbf{S}_{3,3}]$$

составленного из элементов проекций ковариационных матриц на касательное к их средней точке пространство, можно построить устойчивую модель прогнозирования.

В соответствии с *Теоремой 1*, $\phi(\mathbf{X})$ и $\psi(\mathbf{Y})$ являются касательными пространствами к поверхности Римана, а значит между ними существует диффеоморфизм \mathbf{T} , который можно аппроксимировать с помощью модели \mathbf{f} .

Схема метода



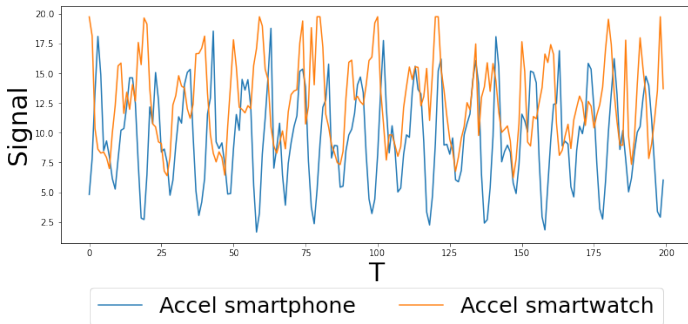
Цели эксперимента

- 1 Проверить качество предсказания с помощью предложенного метода по метрикам: R^2 , MSE
- 2 Изучить зависимость качества прогноза от количества выбранных признаков (QPFS)
- 3 Провести сравнение с другим способом аппроксимации сигналов на основе автокодировщика

Датасет: Smartphone and Smartwatch Activity and Biometrics Dataset

$\{\mathbf{X}, \mathbf{Y}\} = \{x_i, y_i\}_{i=1}^N$ - выборка из сигналов с акселерометра смартфона и смартчасов соответственно.

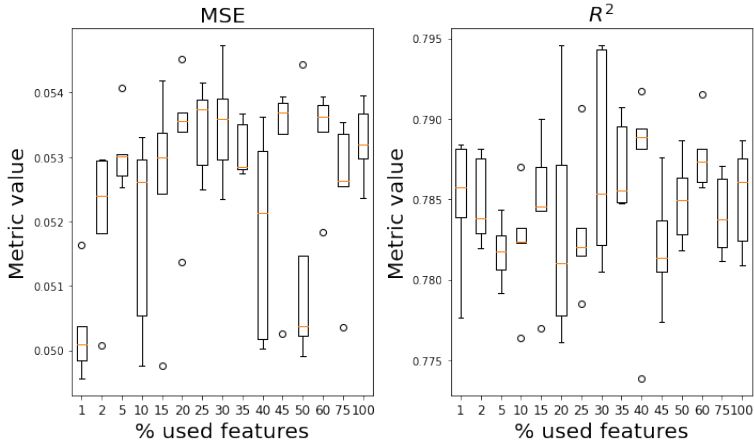
$x_i, y_i \in \mathbb{R}^{3 \times 200}$



1

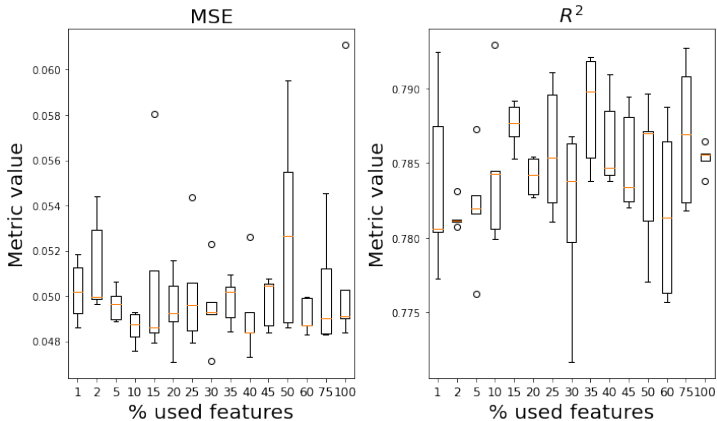
¹Gary M. Weiss, Kenichi Yoneda, and Thaier Hayajneh. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. IEEE Access, 7:133190-133202, Sept. 2019.

Скрытое представление автокодировщика



Высокие показатели коэффициента корреляции. Большой разброс ошибки

Проецирование на многообразии



Использование проекций ковариационных матриц в качестве описания систем позволило достичь более низких показателей ошибки.

- 1 Предложен и теоретически обоснован метод восстановления временных рядов с помощью мультимodelей
- 2 Выполнен численный эксперимент для проверки качества метода
- 3 Показано, что предложенный метод успешно решает задачу корреляционного анализа и показывает высокую устойчивость

Публикации по специальности

- Boginskaya, I.; Safiullin, R.; et al.
Human Angiotensin I-Converting Enzyme Produced by Different Cells: Classification of the SERS Spectra with Linear Discriminant Analysis. Biomedicines 2022, 10, 1389.
- Slipchenko, E.A.; Boginskaya, I.A.; Safiullin, R.R.; et.al.
SERS Sensor for Human Glycated Albumin Direct Assay Based on Machine Learning Methods. Chemosensors 2022, 10, 520.