

Министерство образования и науки Российской Федерации
Московский физико-технический институт (государственный
университет)

Физтех-школа прикладной математики и информатики
Кафедра Интеллектуальных систем

Выпускная квалификационная работа по направлению 030301
Прикладная математика и физика

Пространство параметров модели для
аппроксимация фазовой траектории

Автор:
Христоролюбов Максим Евгеньевич

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

Аннотация

Данная работа посвящена аппроксимации фазовых траекторий временных рядов. Изучается метод аппроксимации, использующий проекции точек фазовой траектории в низкоразмерное пространство, достаточное для восстановления исходной траектории. Предлагается метод построения проецирующего отображения по произвольной модели временного ряда. Изучаются свойства построенного отображения в зависимости от структуры модели временного ряда. Фазовая траектория рассматривается как многообразие, а построенные отображения выступают в роли атласа. Предлагается способ нахождения оптимальной размерности фазовой траектории временного ряда. Для анализа качества аппроксимации проводятся эксперименты на синтетических данных и данных, полученных при помощи акселерометра мобильного устройства.

Ключевые слова: *временной ряд, фазовая траектория, многообразие, атлас, аппроксимация*

Оглавление

1	Введение	5
2	Постановка задачи аппроксимации фазовой траектории	7
2.1	Фазовая траектория ряда как гладкое многообразие	7
2.2	Алгоритм аппроксимации	8
2.3	Оптимальная размерность скрытого пространства	10
2.4	Согласованность атласа	10
3	Модели аппроксимаций физической активности	13
3.1	Анализ сингулярного спектра Гусеница	13
3.2	Автоэнкодер LSTM	14
3.3	Нейронные обыкновенные дифференциальные уравнения	15
3.4	Модель S4	15
3.4.1	Теоретическое обоснование модели	15
3.4.2	Описание модели S4	17
4	Аппроксимация фазовой траектории акселерометра	19
4.1	Синтетические данные и данные акселерометра	19
4.2	Критерии сравнения моделей	20
4.3	Эксперимент	20
5	Заключение	25

Глава 1

Введение

Решается задача аппроксимации фазовой траектории временного ряда. Для аппроксимации точки фазовой траектории используется модель аппроксимации, пространство параметров которой является пространством скрытых переменных. Отображение фазовой траектории в скрытое пространство определяет многообразие, которым является фазовая траектория. Модели аппроксимации образуют атлас.

Цель работы в построении гладко согласованного атласа для фазовой траектории, изучении его свойств и экспериментальной проверке согласованности атласа.

Временной ряд — это ряд наблюдаемых детерминированной динамической системы. **Детерминированный** означает, что в системе нет случайности, даже если она является динамически хаотичной. **Динамический** означает, что значения или состояния системы изменяются с течением времени.

Аттрактор — это пространство состояний, к которым система имеет тенденцию притягиваться с течением времени. Аттрактор можно рассматривать как искривленное пространство, иными словами **многообразие**.

Теневые многообразия (shadow manifolds) — это проекции истинного многообразия системы, которое стоит за наблюдаемыми временного ряда. Теорема Такенса [1] говорит нам, что если ряд \mathbf{s} является измерениями динамической системы, тогда теневое многообразие истинного многообразия M определяется предысторией точки $s_i = X(t_i)$ временного ряда. Точки в этом теневом многообразии M_X биективно соответствуют точкам истинного неизвестного многообразия M . В [2] теневое многообразие называется фазовой траекторией ряда.

Фазовая траектория временного ряда, точки которой являются предысториями каждого момента времени, имеет размерность N по построению. PCA и PLS [?] — это отображение фазовой траектории в низкоразмерное пространство размерности n . PLS проецирует матрицу фазовой траектории исходного временного ряда $\mathbf{X}^{(n)} \in \mathbb{R}^{m \times n}$, где m — число моментов времени временного ряда, а n — размерность фазового пространства и целевую матрицу \mathbf{Y} в скрытое пространство малой размер-

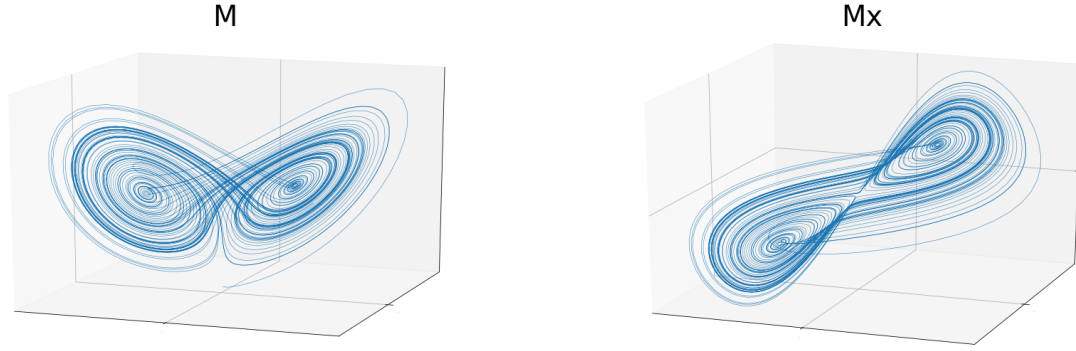


Рис. 1.1: Фазовая траектория системы $M = [X(t), Y(t), Z(t)]$ и теневая фазовая траектория $M_X = [X(t), X(t - \tau), X(t - 2\tau)]$

ностью l ($l < n$). Метод PLS находит в скрытом пространстве матрицы $\mathbf{T}, \mathbf{U} \in \mathbb{R}^{m \times l}$, которые лучше всего описывают исходные матрицы \mathbf{X} и \mathbf{Y} . PLS максимизирует ковариацию между столбцами матриц \mathbf{T} и \mathbf{U} соответственно. Метод PLS соответствует следующей коммутативной диаграмме:

$$\begin{array}{ccc}
 \mathbf{x} \in \mathbb{R}^n & \xrightarrow{\mathbf{f}} & \mathbf{y} \in \mathbb{R}^r \\
 \swarrow \mathbf{P} & & \searrow \mathbf{Q} \\
 & \mathbf{t}, \mathbf{u} \in \mathbb{R}^l & \\
 \nwarrow \mathbf{W} & & \nearrow \mathbf{C}
 \end{array}$$

Рис. 1.2: Диаграмма частичной регрессия наименьших квадратов

Данному подходу соответствует авторегрессионная модель [3], сингулярное спектральное разложение (SSA) [4]. В качестве низкоразмерного пространства в них выступает пространство параметров модели временного ряда.

Глава 2

Постановка задачи аппроксимации фазовой траектории

2.1 Фазовая траектория ряда как гладкое многообразие

Определение 1. *Временным рядом называется функция дискретного аргумента $\mathbf{s}(t)$, сопоставляющая отсчетам времени $t_i \in \mathcal{T}$ вектор значений измеряемых переменных $\mathbf{s}(t_i) = s_i \in \mathbb{S} = \mathbb{R}^d$.*

В работе рассматриваются временные ряды с $d = 1$, в которых измерения проведены в одни и те же моменты времени $t \in \mathcal{T}$, такие, что $\Delta t = t_{i+1} - t_i$ постоянна.

Определение 2. *Предысторией длины N для момента времени t_i временного ряда $\mathbf{s} = [s_0 \dots s_m] \in \mathbb{R}^1$ является $\mathbf{x}_i^{(N)} = [s_{i-N+1} \dots s_i] \in \mathbb{X} = \mathbb{R}^N$.*

Сопоставление моментам временного ряда их предыстории осуществляется отображением $\eta : \eta(t_i, s) = x_i$, в дальнейшем зависимость от s опускается при записи для простоты.

Доступ к фазовой траектории динамической системы отсутствует, поэтому вместо нее работают с фазовой траекторией ряда. Для ее построения каждой точке s_i временного ряда ставится в соответствие его предыстория

$$\mathbf{x}_i^{(N)} = \eta(s_i).$$

В данном случае $\mathbf{x}_i^{(N)}$ — точки фазовой траектории.

Гладкое многообразие — множество с выделенным классом попарно согласованных локальных параметризаций, где согласованность означает, что две параметризации переводятся друг в друга диффеоморфизмом.

Определение 3. [5] Гладким n -мерным многообразием M называется множество, для которого задана система подмножеств X_i и взаимно однозначные отображения на них $\phi_i : W_i \rightarrow X_i$ открытых подмножеств W_i аффинного пространства \mathbb{R}^n , причем

$$1) M = \cup X_i,$$

2) Для каждой пары ϕ_i, ϕ_j прообразы пересечения $X_i \cap X_j$ — множества

$$W_{ij} = \phi_i^{-1}(X_i \cap X_j) \text{ и } W_{ji} = \phi_j^{-1}(X_i \cap X_j)$$

являются открытыми подмножествами в \mathbb{R}^n ,

3) $\phi_{ij} = \phi_j^{-1} \phi_i$ есть диффеоморфизм

$$W_{ij} = \phi_i^{-1}(X_i \cap X_j) \text{ на } W_{ji} = \phi_j^{-1}(X_i \cap X_j).$$

Определение 4. Взаимнооднозначное отображение: $\phi : W \rightarrow X$, где W — область в \mathbb{R}^n , $X \subset M$, называется в общем случае локальной параметризацией, также картой многообразия M или локальной координатной системой. Две карты называются гладко согласованными, если для них выполнено условие 3) определения 3.

Размерность многообразия определяется по размерности евклидова пространства, с которым оно локально сходно, а именно размерностью W_i , которая равна n

Определение 5. Совокупность карт ϕ_i называется атласом, если области X_i покрывают M . Если выполнены три условия определения 3, то говорят, что данный атлас является гладко согласованным и определяет в M структуру гладкого многообразия.

Фазовая траектория является многообразием. Действительно, пусть имеется динамическая система, тогда всевозможные точки фазовой траектории образуют многообразие. В данной работе в роли динамической системы выступает ходьба человека, а фазовая траектория строится по показаниям акселерометра на теле. Более подробно фазовая траектория как многообразие рассмотрена в следующих разделах.

2.2 Алгоритм аппроксимации

Обозначим

$$\mathbf{x} = [x_i]_{i=0}^m \text{ — точка фазовой траектории.}$$

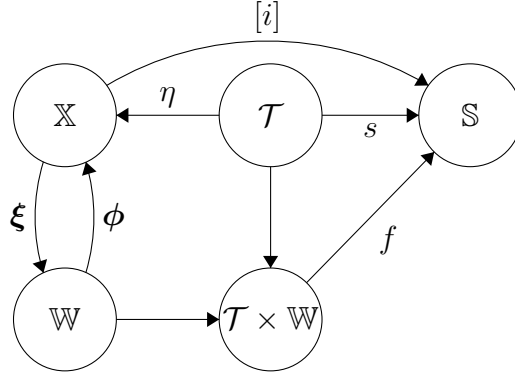
Определение 6. Модель временного ряда — это параметризованная скалярная функция $f(t, \mathbf{w})$. Существует значения параметра \mathbf{w}_0 оптимальное в смысле минимизации функции ошибки $L(\mathbf{x}, f(t, \mathbf{w}))$, такое что ряд $[f(t_i, \mathbf{w}_0)]_{i=0}^m$ аппроксимирует исходный ряд $[x_i]_{i=0}^m$.

Тогда отображение в скрытое пространство \mathbb{W} — это нахождение оптимального значения параметра \mathbf{w}_0 для модели $f(t, \mathbf{w})$ и функции ошибки $L(\mathbf{x}, f(t, \mathbf{w}))$.

Определение 7. Отображение $\xi : \mathbb{X} \hookrightarrow \mathbb{W}$ объектов выборки $\mathbf{X} \in \mathbb{R}^N$ называется вложением \mathbf{X} в скрытое пространство $\mathbb{W} = \mathbb{R}^n$

Определение 8. Аппроксимацией точки \mathbf{x} фазовой траектории называется композиция $\xi \circ \phi$ вложения точки фазовой траектории в скрытое пространство $\xi : \mathbb{X} \hookrightarrow \mathbb{W}$ и восстановления $\phi : \mathbb{W} \hookrightarrow \mathbb{X}$.

Общую структуру аппроксимации иллюстрирует диаграмма:



Схожие обозначения в этом и предыдущем разделе выбраны не случайно. Точки \mathbf{x}_i фазовой траектории, каждая из своей окрестности X_i , лежат на многообразии

$$M = \cup X_i.$$

Доступ ко всему многообразию отсутствует, а имеется конечный набор точек

$$\mathbf{X} = [\mathbf{x}_0, \dots, \mathbf{x}_m]^\top.$$

Каждая точка \mathbf{w}_i скрытого пространства восстанавливает точку фазовой траектории $\hat{\mathbf{x}}_i = [f(t_j, \mathbf{w}_i)]_{j=0}^N$. Тогда из соображений непрерывности данная функция восстановления обобщается на окрестность W_i точки w_i и строится $\phi_i : W_i \hookrightarrow X_i$ так, что для всех $\mathbf{w} \in W_i$ выполнено

$$\phi_i(\mathbf{w}) = \hat{\mathbf{x}}_i$$

.

Таким образом из алгоритма аппроксимации естественным образом возникает система отображений

$$\phi_i : W_i \rightarrow X_i \text{ и обратных отображений } \phi_i^{-1} = \xi_i : X_i \hookrightarrow W_i$$

2.3 Оптимальная размерность скрытого пространства

Размерность N фазового пространства, содержащего предысторию точек временного ряда, велика по построению. При проекции размерность пространства понижается с помощью отображения точек фазовой траектории $\mathbf{x} \in \mathbb{R}^N$ в $\mathbf{w} \in \mathbb{R}^n$. Вводится понятие оптимальной размерности скрытого пространства.

Поскольку скрытое пространство \mathbb{W} — это пространство параметров моделей, аппроксимирующих точки фазовой траектории, то проблема выбора оптимальной размерности совпадает с проблемой выбора оптимальной сложности модели аппроксимации временного ряда. Избыточная сложность модели приведет к переобучению и тому, что в скрытом пространстве существует подпространство, все точки которого аппроксимируют единственную точку фазовой траектории, что означает, что требование взаимнооднозначности ϕ не выполнено. Недостаточная сложность модели приведет к потере информации, что означает, что различные точки фазовой траектории проецируются в одну точку скрытого пространства и не могут быть корректно восстановлены, что опять противоречит взаимнооднозначности. Поэтому, если модель аппроксимации корректна, то существует оптимальная размерность пространства n , при которой выполняются требование взаимнооднозначности отображения ϕ .

Оптимальная размерность скрытого пространства задается размерностью многообразия, представляющего фазовую траекторию. Таким образом, если удалось построить гладко согласованный атлас, то размерность W_i является оптимальным размером скрытого пространства.

2.4 Согласованность атласа

Цель данной работы в составлении из различных моделей атласа и проверки является ли атлас гладко согласованным. Для этого требуется, чтобы ϕ было взаимнооднозначным, а $\xi \circ \phi$ диффеоморфным.

Очевидно, что ϕ не является взаимнооднозначным, так как одному восстановленному $\hat{\mathbf{x}}_i$ соответствуют все \mathbf{w} из окрестности W_i . Однако, это не является проблемой, поскольку на практике различие между очень близкими друг к другу точками фазовой траектории обусловлено случайным шумом, по факту они представляют одну и ту же точку фазовой траектории не зашумленной динамической системы. Поэтому требование однозначности для не зашумленной динамической системы выполнено, если размерность скрытого пространства совпадает с размерностью многообразия фазовой траектории.

Для того, чтобы проверить $\xi \circ \phi$ на диффеоморфность следует изучить дифференцируемость ξ и $\phi = [f(t_j, \mathbf{w})]_{j=0}^N$. В данной работе все модели $f(t, \mathbf{w})$ дифферен-

цируемы по своим параметрам

$$\frac{\partial f(\cdot, \mathbf{w})}{\partial w} = \frac{\partial \phi(\mathbf{w})}{\partial w}.$$

Чтобы ξ было дифференцируемо требуется, чтобы параметры \mathbf{w} модели, полученные как проекция точек \mathbf{x} фазового пространства \mathbb{X} , были дифференцируемы относительно исходных точек $\frac{\partial \mathbf{w}}{\partial \mathbf{x}} = \frac{\partial \xi(\mathbf{x})}{\partial \mathbf{x}}$.

Для модели временного ряда $f(t, \mathbf{w})$ временного ряда вложение ξ строится с использованием минимизации функции ошибки L . Эта техника используется в авторегрессионных моделях AR, ARIMA [6, 7], в моделях Фурье и при аппроксимации сплайнами [8], SSA Гусеница.

$$\mathbf{w} = \arg \min_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w})).$$

Так как каждой точке \mathbf{x} фазовой траектории соответствует набор параметров модели \mathbf{w} , то данное соответствие задает функцию $\xi(\mathbf{x}) = \mathbf{w}_{\mathbf{x}}$.

Для того, чтобы модели образовывали согласованный атлас, они должны быть гладко дифференцируемой, отсюда возникает вопрос о дифференцируемости ξ .

Теорема 1. Пусть дана функция

$$\xi(\mathbf{x}) = \arg \min_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w})).$$

Если $L(\mathbf{x}, \phi(\mathbf{w}))$ — выпуклая функция, а $\phi(\mathbf{w})$ — линейная функция, тогда $\xi(\mathbf{x})$ — гладко дифференцируемая функция.

Доказательство.

Согласно необходимому условию минимума функции L :

$$\partial_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} = 0.$$

Дифференцируя равенство по \mathbf{x} , получим:

$$\partial_{\mathbf{x}} \partial_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} + \xi'(\mathbf{x}) \cdot \partial_{\mathbf{w}}^2 L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} = 0.$$

Отсюда находится явная формула для $\frac{\partial \mathbf{w}}{\partial \mathbf{x}} = \xi'(\mathbf{x})$, но для этого требуется, чтобы второй множитель не был равен 0:

$$\partial_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w})) = \phi'(\mathbf{w}) \cdot \partial_{\phi} L(\mathbf{x}, \phi(\mathbf{w})),$$

$$\partial_{\mathbf{x}} \partial_{\mathbf{w}} L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} = \phi'(\mathbf{w}) \cdot \partial_{\mathbf{x}} \partial_{\phi} L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})},$$

$$\partial_{\mathbf{w}}^2 L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} = \phi''(\mathbf{w}) \cdot \partial_{\phi} L(\mathbf{x}, \phi(\mathbf{w})) + (\phi'(\mathbf{w}))^2 \cdot \partial_{\phi}^2 L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})}.$$

Произведение

$$(\phi')^2 \cdot \partial_{\phi}^2 L(\mathbf{x}, \phi) > 0,$$

так как L — выпуклая функция. Если модель f линейная, то $\phi''(\mathbf{w}) = 0$ и множитель $\partial_{\mathbf{w}}^2 L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} \neq 0$. Кроме того $\xi'(\mathbf{x})$ непрерывна.

Следовательно, при заданных условиях отображение ξ гладко дифференцируемо.

Таким образом, в модели ARIMA ξ является гладко дифференцируемой в соответствии с доказанной теоремой, так как L — квадратичная функция ошибки, а f линейно зависит от своих параметров.

В модели SSA Гусеница ξ вычисляется с помощью матричных операций и сингулярного разложения, что означает дифференцируемость ξ .

Доказанная теорема является строгим доказательством дифференцируемости ξ . Все требуемые в теореме условия нужны, чтобы знаменатель в выражении для $\xi'(x)$ не был равен 0 ни в каких точках. Однако на практике, если знаменатель будет обращаться в 0 на множестве точек \mathbf{x} нулевой меры, то отсутствие дифференцируемости в этих точках заметно не будет. Множество корней уравнения

$$\partial_{\mathbf{w}}^2 L(\mathbf{x}, \phi(\mathbf{w}))|_{\mathbf{w}=\xi(\mathbf{x})} = 0$$

является множеством меры 0, поэтому построенное данным образом отображение не отличимо от гладкого отображения. Следовательно, согласованность атласа определяется исключительно непрерывностью $\xi \circ \phi$, из чего следует взаимнооднозначность ϕ .

Глава 3

Модели аппроксимаций физической активности

В данной работе сравнивается качество аппроксимации нескольких моделей временных рядов и исследуется возможность составления их гладко согласованного атласа.

3.1 Анализ сингулярного спектра Гусеница

Модель Сингулярного спектра (SSA) представляет сегмент временного ряда в виде суммы интерпретируемых компонент. Точке фазовой траектории \mathbf{x}_i^N ставится в соответствие его траекторная матрица Ганкеля $\mathbf{X}_i^N \in R^{N/2 \times N/2+1}$:

$$\mathbf{X}_i^{(N)} = \begin{pmatrix} x_{i-N+1}^{(N)} & x_{i-N+2}^{(N)} & \cdots & x_{i-\frac{N}{2}+1}^{(N)} \\ x_{i-N+2}^{(N)} & x_{i-N+3}^{(N)} & \cdots & x_{i-\frac{N}{2}+2}^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i-\frac{N}{2}}^{(N)} & x_{i-\frac{N}{2}+1}^{(N)} & \vdots & x_i^{(N)} \end{pmatrix},$$

где N — выбирается равным 2 характерным периодам ряда.

Сингулярное разложение матрицы $\mathbf{X}_i^{(N)}$:

$$\mathbf{X}_i^{(N)} = \mathbf{U}(\mathbf{h})\mathbf{V} = \sum_{j=1}^{N/2} h_j \mathbf{u}_j \mathbf{v}_j^T, \quad (3.1)$$

где $h_1 \dots h_T$ — сингулярные числа матрицы $\mathbf{X}_i^{(N)}$, наибольшие из которых будут компонентами проекции в низкоразмерное пространство $\mathbf{w} = \xi(\mathbf{x}) = \mathbf{h}$.

Чтобы восстановить исходную точку фазовой траектории матрицы $\mathbf{U}(\mathbf{h})\mathbf{V}$ усредняются по антидиагоналям, каждая из полученных усредненных матриц является Ганкелевой матрицей для сегмента временного ряда $\hat{\mathbf{x}}_k$. Восстановленная точка фазовой траектории находится как сумма этих сегментов:

$$\hat{\mathbf{x}} = \sum_{k=1}^K \hat{\mathbf{x}}_k.$$

Собственные числа матрицы гладко зависят от самой матрицы [9], значит, сингулярные числа \mathbf{h} гладко зависят от матрицы \mathbf{X} , и от \mathbf{x} , т. е. ξ — гладкое отображение в данном случае. Функция восстановления временного ряда ϕ является композицией суммирования и частного, поэтому тоже является гладкой.

3.2 Автоэнкодер LSTM

LSTM (долгая краткосрочная память) — тип рекуррентной нейронной сети, способный обучаться долгосрочным зависимостям. LSTM были представлены в [10], впоследствии усовершенствованы и популяризированы другими исследователями, хорошо справляются со многими задачами и до сих пор широко применяются.

Все рекуррентные нейронные сети имеют форму цепочки повторяющихся модулей нейронной сети.

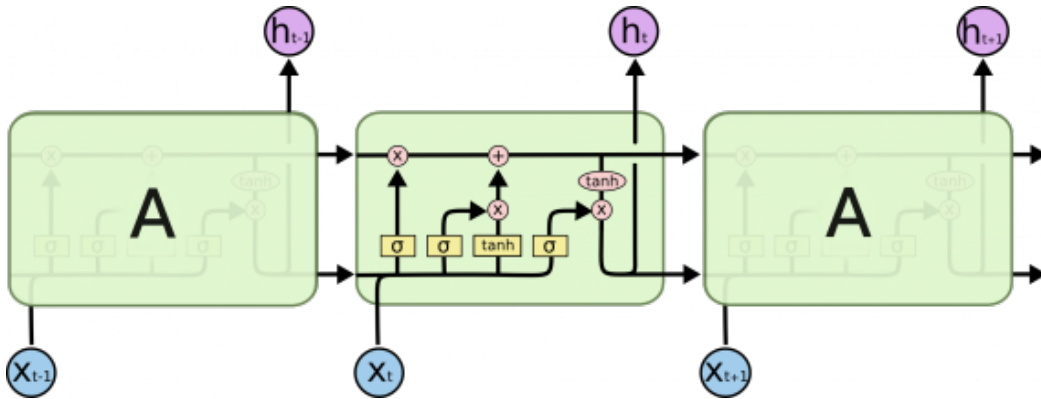


Рис. 3.1: Структура слоя LSTM [10]

$$\begin{aligned} \mathbf{f}_t &= \sigma_g(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \\ \mathbf{i}_t &= \sigma_g(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \\ \mathbf{o}_t &= \sigma_g(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \\ \mathbf{c}_t &= \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \sigma_c(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \\ \mathbf{h}_t &= \mathbf{o}_t \circ \sigma_h(\mathbf{c}_t) \end{aligned}$$

3.3 Нейронные обыкновенные дифференциальные уравнения

В нейронном ОДУ предполагается, что временной ряд — это измерения непрерывной функции, которая удовлетворяет ОДУ:

$$\frac{ds}{dt} = g(s, w)$$

Цель подхода состоит не в том, чтобы восстановить значения ряда s , а в восстановлении производной $\frac{ds}{dt}$, которая определяет поведение временного ряда. Данный подход позволяет более качественно восстанавливать исходный ряд, если ряд представляет измерения физического явления, удовлетворяющего дифференциальному уравнению. Предсказание данной модели получают решением дифференциального уравнения с начальными условиями, в котором правая часть задается нейронной сетью $g(s, w)$. Это означает, что один прямой проход дает нам всю траекторию, в отличие от рекуррентных сетей, где каждый прямой проход через модель дает одно предсказание во времени. Решение ОДУ имеет вид:

$$s(t) = s(t_0) + \int_{t_0}^t g(s(t), w) dt.$$

Нейросеть $g(s, w)$ внутри ОДУ обучается стандартными градиентными методами, основанными на обратном распространении ошибки.

3.4 Модель S4

3.4.1 Теоретическое обоснование модели

Данная модель основывается на проекции фазовой траектории ряда в низкоразмерное пространство. Предлагается ввести на предыстории $\mathbf{x}_t^{(\xi)}$ меру $\mu^{(\xi)}(t)$, например,

$$\mu^{(\xi)}(t) = \frac{1}{\xi} \mathbb{I}_{[t-\xi, t]}(t),$$

тогда параметры модели в каждый момент времени t зависят только о предыстории $[t - \xi, t]$. Данную меру называют Транслированной мерой Лежандра (translated Legendre measures LegT). В статье [11] используется равномерная мера для всей истории

$$\mu^{(\xi)}(t) = \frac{1}{\xi} \mathbb{I}_{[0, \xi]}(t) \text{ (LegS)}.$$

Выбранная мера порождает на пространстве функций скалярное произведение

$$\langle f, g \rangle_\mu = \int f(t)g(t)d\mu(t)$$

и индуцирует гильбертовое пространство с нормой

$$\|f\|_{L_2(\mu)} = \langle f, f \rangle_\mu^{\frac{1}{2}}$$

Предыстория \mathbf{x}_ξ проецируется на n -размерное подпространство, при этом вложением фазовой траектории является разложение предыстории в ортогональном базисе.

Для скрытого состояния \mathbf{w} выполнено:

$$\mathbf{w} = \arg \min_{\mathbf{w}} \|\hat{\mathbf{x}} - \mathbf{x}\|_{L_2(\mu)} = \arg \min_{\mathbf{w}} \int (\hat{s}(t) - s(t))^2 d\mu^{(\xi)}(t)$$

Для конкретного момента времени ξ , скрытое состояние $\mathbf{w}(\xi)$ находится с помощью скалярного произведения. Но функция $\mathbf{w}(t)$ является решением ОДУ $\frac{d}{dt}\mathbf{w}(t) = \mathbf{A}(t)\mathbf{w}(t) + \mathbf{B}(t)s(t)$, где $\mathbf{A}(t) \in \mathbb{R}^{h \times h}$, $\mathbf{B}(t) \in \mathbb{R}^{h \times 1}$. Таким образом, данное ОДУ сводится к рекуррентному соотношению и вычислять динамически.

В статье [11] в качестве базиса $\mathbf{w}(t)$, линейной оболочкой которого приближается предыстория, являются многочлены Лежандра $P_n(t)$. Известно, что они ортогональны относительно скалярного произведения, индуцированного мерой $\mathbf{1}_{[-1,1]}$:

$$\frac{2n+1}{2} \int_{-1}^1 P_n(t)P_m(t)dt = \delta_{nm}$$

Ортогональный базис $g^{(\tau)}(t)$ относительно $\mu^{(\tau)}(t) = \frac{1}{\tau}\mathbb{I}_{[0,\tau]}$ получается с помощью замены переменных и нормировки:

$$g_n = (2n+1)^{\frac{1}{2}} P_n\left(\frac{2t}{\tau} - 1\right)$$

Алгоритм нахождения матрицы \mathbf{A} (матрицы НРРО [12]):

- 1) продифференцировать равенство $\mathbf{w}_n(\xi) = \langle x, g^{(\xi)} \rangle_{\mu^{(\xi)}}$ по переменной ξ с помощью формулы Лейбница,
- 2) привести полученное соотношение к виду $\frac{d}{d\xi}\mathbf{w}(\xi) = \mathbf{A}\mathbf{w}(\xi) + \mathbf{B}s(\xi)$,
- 3) дискретизировать ОДУ и получить рекуррентное соотношение.

3.4.2 Описание модели S4

Модель S4, используемая в [11], задается системой уравнений, в которой $s(t_0), y(t_0) \in \mathbb{R}$, $\mathbf{w}(t_0) \in \mathbb{R}^h$, выход модели y :

$$\mathbf{w}'(t) = \mathbf{A}\mathbf{w}(t) + \mathbf{B}s(t),$$

$$y(t) = \mathbf{C}\mathbf{w}(t) + \mathbf{D}s(t).$$

Матрица \mathbf{A} инициализируется следующим образом:

$$A_{nk} = - \begin{cases} (2n+1)^{\frac{1}{2}}(2k+1)^{\frac{1}{2}} & \text{при } n > k, \\ n+1 & \text{при } n = k, \\ 0 & \text{при } n < k. \end{cases}$$

Матрица представляется в виде

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^* - \mathbf{P}\mathbf{Q}^*,$$

где $\mathbf{V} \in \mathbb{C}^{h \times h}$, $\mathbf{\Lambda}$ диагональная и $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{h \times 1}$. Обучаемые параметры модели - диагональная матрица $\mathbf{\Lambda}$ и векторы $\mathbf{P}, \mathbf{Q}, \mathbf{B}, \mathbf{C}, \mathbf{D}$.

Дискретное представление получается из непрерывного с помощью дискретизация непрерывного времени по правилу $s_k = s(k\Delta t)$:

$$\mathbf{w}_k = \bar{\mathbf{A}}\mathbf{w}_{k-1} + \bar{\mathbf{B}}s_k, \quad y_k = \bar{\mathbf{C}}\mathbf{w}_k + \bar{\mathbf{D}}s_k,$$

$$\bar{\mathbf{A}} = (\mathbf{I} - \frac{\Delta t}{2}\mathbf{A})^{-1}(\mathbf{I} + \frac{\Delta t}{2}\mathbf{A}), \quad \bar{\mathbf{B}} = (\mathbf{I} - \frac{\Delta t}{2}\mathbf{A})^{-1}\Delta\mathbf{B}, \quad \bar{\mathbf{C}} = \mathbf{C}, \quad \bar{\mathbf{D}} = \mathbf{D}.$$

Реккурентное представление модели является RNN: \mathbf{w}_k — это скрытое состояние матрицы перехода $\bar{\mathbf{A}}$.

Пусть $\mathbf{w}_{-1} = 0$, тогда из рекуррентного соотношения:

$$\mathbf{w}_0 = \bar{\mathbf{B}}s_0, \quad y_0 = \bar{\mathbf{C}}\bar{\mathbf{B}}s_0,$$

$$\mathbf{w}_1 = \bar{\mathbf{A}}\bar{\mathbf{B}}s_0 + \bar{\mathbf{B}}s_1, \quad y_1 = \bar{\mathbf{C}}\bar{\mathbf{A}}\bar{\mathbf{B}}s_0 + \bar{\mathbf{C}}\bar{\mathbf{B}}s_1$$

$$\mathbf{w}_2 = \bar{\mathbf{A}}^2\bar{\mathbf{B}}s_0 + \bar{\mathbf{A}}\bar{\mathbf{B}}s_1 + \bar{\mathbf{B}}s_2, \quad y_2 = \bar{\mathbf{C}}\bar{\mathbf{A}}^2\bar{\mathbf{B}}s_0 + \bar{\mathbf{C}}\bar{\mathbf{A}}\bar{\mathbf{B}}s_1 + \bar{\mathbf{C}}\bar{\mathbf{B}}s_2$$

$$y_k = \bar{\mathbf{C}}\bar{\mathbf{A}}^k\bar{\mathbf{B}}s_0 + \bar{\mathbf{C}}\bar{\mathbf{A}}^{k-1}\bar{\mathbf{B}}s_1 + \dots + \bar{\mathbf{C}}\bar{\mathbf{B}}s_k$$

$$y_k = \bar{\mathbf{K}} * s$$

Таким образом, $y_k = \bar{\mathbf{K}} * s$ — это сверточное представление модели, которое эффективно вычисляется, если известно $\bar{\mathbf{K}}$. Существует алгоритм быстрого нахождения $\bar{\mathbf{K}}$ в случае данного специфического устройства матрицы $\bar{\mathbf{A}}$.

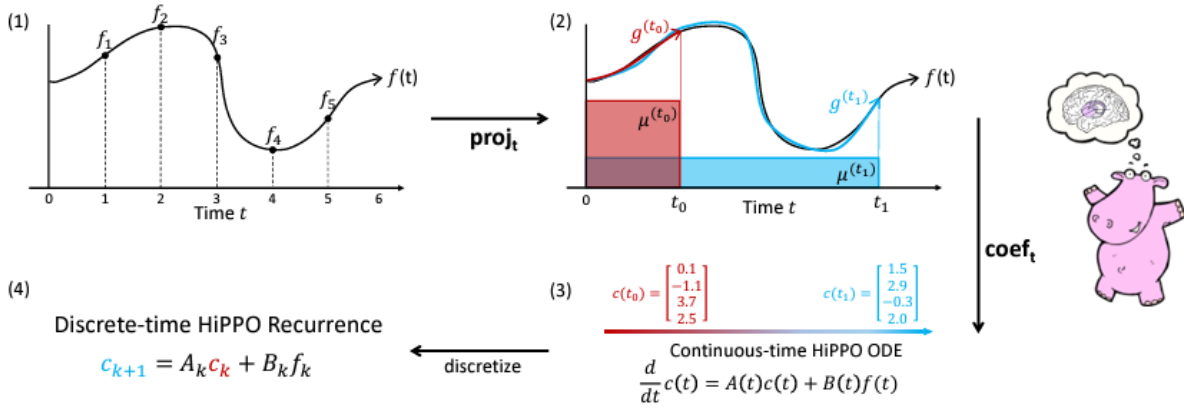


Рис. 3.2: диаграмма алгоритма HiPPO

Теорема 2. Непрерывное и дискретное уравнение скрытого состояния системы для HiPPO-LegS:

$$\frac{d}{dt}\mathbf{w}(t) = -\frac{1}{t}\mathbf{A}\mathbf{w}(t) + \frac{1}{t}\mathbf{B}s(t), \quad \mathbf{w}_{k+1} = \left(1 - \frac{\mathbf{A}}{k}\right)\mathbf{w}_k + \frac{1}{k}\mathbf{B}s_k$$

$$A_{nk} = - \begin{cases} (2n+1)^{\frac{1}{2}}(2k+1)^{\frac{1}{2}} & \text{при } n > k, \\ n+1 & \text{при } n = k, \\ 0 & \text{при } n < k, \end{cases} \quad B_n = (2n+1)^{\frac{1}{2}}$$

Оператор HiPPO обладает благоприятными теоретическими свойствами: он инвариантен по отношению к частоте дискретизации временного ряда (коэффициенты аппроксимации $\mathbf{w}(t)$ не меняются при изменении масштаба), быстр в вычислениях, имеет ограниченные градиенты и ошибку аппроксимации.

Глава 4

Аппроксимация фазовой траектории акселерометра

Выбирается размерность скрытого пространства произвольным образом, заведомо превосходящая оптимальную размерность. После каждая точка \mathbf{x}_i фазовой траектории в выборке отображается в скрытое пространство, после чего декодируется обратно в $\hat{\mathbf{x}}_i$.

4.1 Синтетические данные и данные акселерометра

Синтетический данные сгенерированы по формуле

$$x_i(t) = \sin t + a_i \sin \frac{t}{2} + N\left(0, \frac{1}{5}\right) \sin \frac{t}{9} + N\left(0, \frac{1}{10}\right)$$

Здесь первое слагаемое отвечает за главную составляющую динамической системы (ходьба), второе за исследуемый признак каждой точки фазовой траектории (вес рюкзака), третье слагаемое отвечает за неучтенные факторы, влияющие на динамическую систему, а четвертое — случайный шум. Рассматриваются разные формулы для a_i : $a_i = \frac{i+10}{100}$, $a_i = \frac{1}{3} + \frac{1}{100-i}$

Реальные данные — это измерения акселерометра, встроенных в мобильное устройство, хранящегося в переднем кармане брюк участника. Временные ряды содержат значения ускорения человека для оси Z акселерометра. Частота дискретизации составляет 50 Гц. Данные собраны с одного и того же человека, идущего по прямой с рюкзаком различного веса.

Во всех выорках объекты отсортированы по исследуемому признаку (a_i и вес рюкзака).

4.2 Критерии сравнения моделей

В работе рассматриваются модели аппроксимации фазовой траектории системы и сравниваются по следующим критериям:

1) **Точность** аппроксимации. В качестве метрики используется отклонение, а именно **L2 норма разности** исходного и восстановленного временного ряда:

$$\sum_{i=1}^m (\hat{x}_i - x_i)^2$$

2) **Оптимальность** (сложность модели), в качестве метрики - размерность скрытого пространства.

Поскольку для используемых на практике функций для аппроксимации временных свойства липшицевости и гладкости эквивалентны, то вместо проверки на гладкость атласа исследуется его липшицевость.

3) **Липшицевость**. Исследуется устойчивость в том смысле, что если исходный ряд претерпевает небольшие изменения (в смысле L2 нормы), то восстановленный ряд не должен претерпевать больших изменений (в смысле L2 нормы). Этот критерий можно так же воспринимать как **устойчивость**.

Предлагается следующим образом оценивать устойчивость:

- а) Получить аппроксимацию $\hat{\mathbf{x}}(t)$ точки $\mathbf{x}(t)$.
- б) Поскольку объекты выборки отсортированы по исследуемому признаку и близки друг к другу, то достаточно вычислить метрики

$$\text{StabError}(\mathbf{x}_i, \mathbf{x}_{i+1}) = \frac{\|\mathbf{w}_i - \mathbf{w}_{i+1}\|}{\|\mathbf{x}_i - \mathbf{x}_{i+1}\|}$$

$$\text{MeanStabError}(\mathbf{x}) = \frac{1}{\text{len}(\mathbf{X}) - 1} \sum_i \text{StabError}(\mathbf{x}_i)$$

$$\text{MaxStabError}(\mathbf{x}) = \max_i (\text{StabError}(\mathbf{x}_i))$$

Если липшицевость, а как следствие дифференцируемость отсутствует, значит размерность скрытого пространства избыточна. Будем понижать размерность, пока аппроксимация не будет липшицевой.

4.3 Эксперимент

Пример того, как SSA аппроксимирует ряд

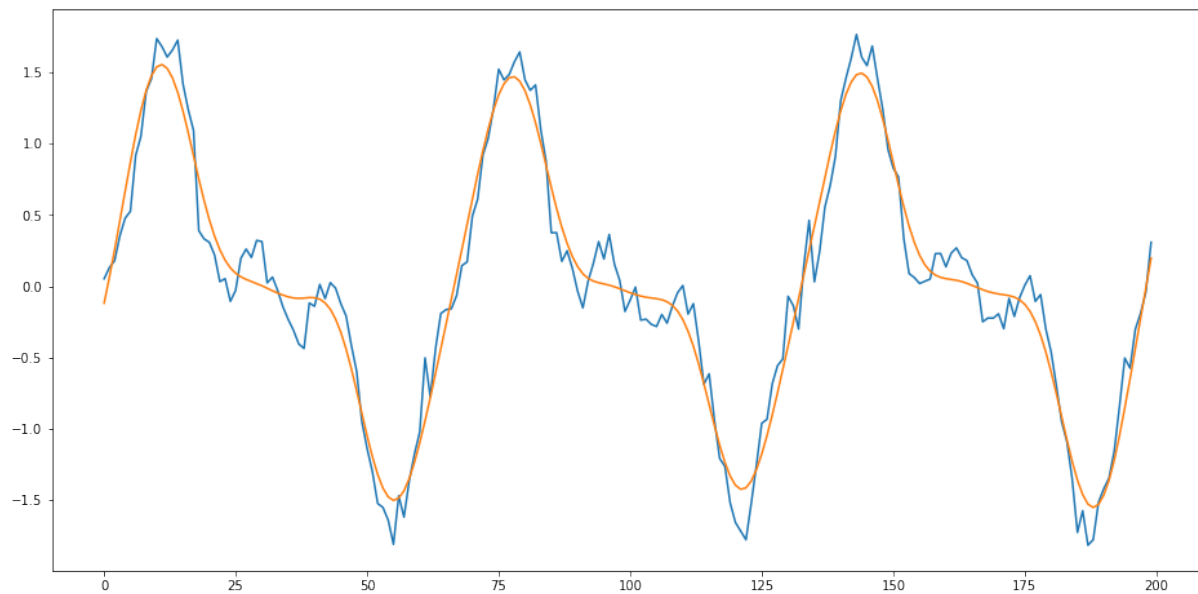


Рис. 4.1: Аппроксимация синтетического ряда

Значения параметров в низкоразмерном пространстве, цветом показано значение исследуемого признака:

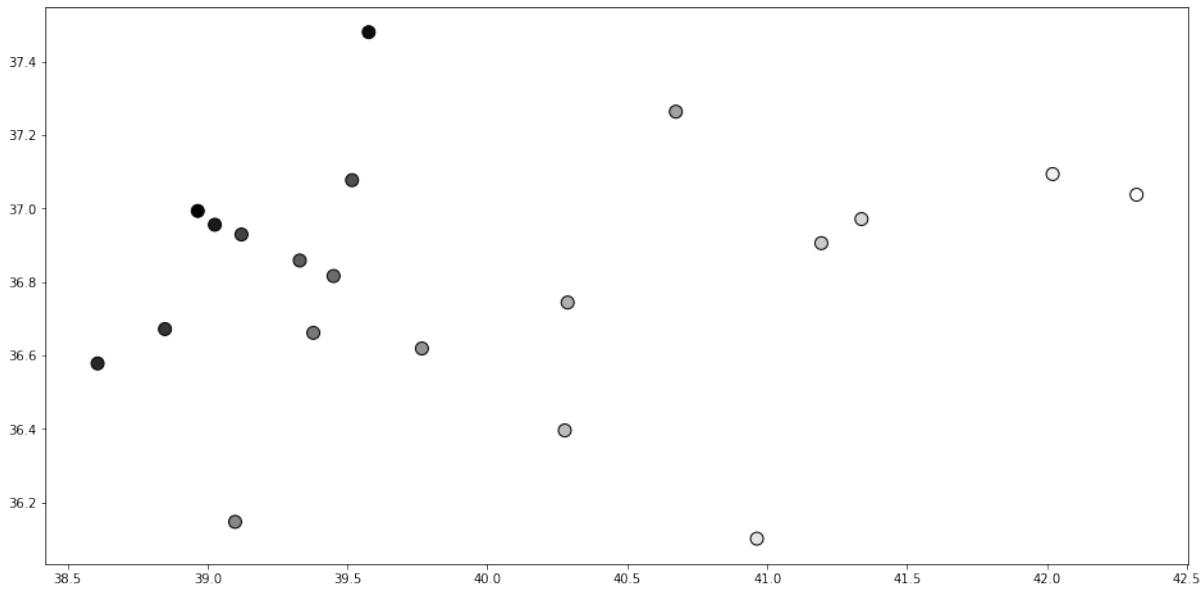


Рис. 4.2: Параметры SSA для разных рядов

Чтобы убедиться в липшицевости отображения ξ визуализируем фазовую траекторию ряда и траекторию в пространстве параметров модели. Красным кругом указаны точки фазовой траектории и область в пространстве параметров, куда они отображаются.

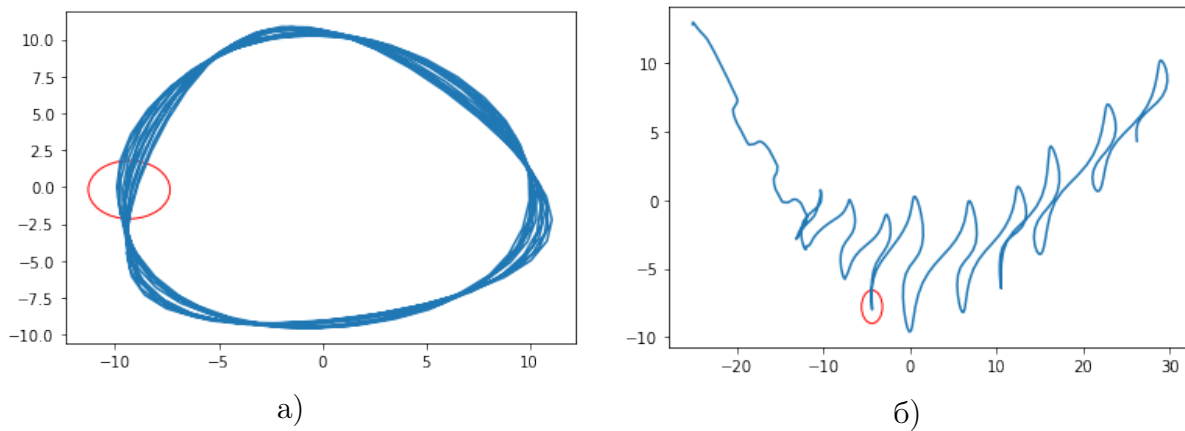


Рис. 4.3: Фазовая траектория и траектория в пространстве параметров модели

Результаты моделей аппроксимации на синтетической выборке:

Модель	Std	Размерность СП	MeanStabError	MaxStabError	max L
SSA	2.12	2	2.06	3.98	0.35
LSTM	4.21	2	1.80	3.63	0.51
S4	3.93	2	1.20	3.52	0.44

Результаты моделей аппроксимации на данных акселерометра мобильного устройства:

Модель	Std	Размерность СП	MeanStabError	MaxStabError	max L
SSA	30.0	10	1.80	5.22	0.64
LSTM	44.2	10	3.08	7,88	1,04
S4	33.1	10	2,14	6,33	0,98

Глава 5

Заключение

В работе показано, что фазовая траектория временного ряда является многообразием и как все модели аппроксимации косвенно используют этот факт. Проверена гипотеза о том, что по дифференцируемости отображения в скрытое пространство возможно определить оптимальный размер скрытого пространства. Проведено экспериментальное сравнение моделей аппроксимации по описанным критериям.

Литература

- [1] *Takens, F.* Detecting strange attractors in turbulence / F. Takens // Dynamical Systems and Turbulence, Lecture Notes in Mathematics. — 1981.
- [2] *Исаченко Р.В., Стрижов В.В.* Снижение размерности пространства в задачах декодирования сигналов / Стрижов В.В. Исаченко Р.В. — 2021.
- [3] *Lukashin, Y.P.* Adaptive methods of short-term forecasting of time series / Y.P. Lukashin // Finance and statistics. — 2003.
- [4] *Hassani, H.* Singular spectrum analysis: methodology and comparison. Journal of Data Science / H. Hassani. — 2007.
- [5] *Чернавский, А.В.* / А.В. Чернавский // Часть первая. Многообразия. — 2010.
- [6] *Agrawal, Ratnadip Adhikari R. K.* An Introductory Study on Time Series Modeling and Forecasting / Ratnadip Adhikari R. K. Agrawal. — 2013.
- [7] *Н. В. Артамонов Е. А. Ивин, А. Н. Курбацкий Д. Фантацини* / А. Н. Курбацкий Д. Фантацини Н. В. Артамонов, Е. А. Ивин // Введение в анализ временных рядов. — 2021.
- [8] *Д. А. Аникеев Г. О. Пенкин, В. В. Стрижов.* Классификация физической активности человека с помощью локальных аппроксимирующих моделей / В. В. Стрижов Д. А. Аникеев, Г. О. Пенкин // Информ. и её применение. — 2019.
- [9] *Andreas Kriegl Peter W. Michor, Armin Rainer.* Denjoy-Carleman differentiable perturbation of polynomials and unbounded operators / Armin Rainer Andreas Kriegl, Peter W. Michor. — 2009.
- [10] *Sepp Hochreiter Jurgen Schmidhuberl, Peter W. Michor Armin Rainer.* Long Shortterm Memory, Neural Computation / Peter W. Michor Armin Rainer Sepp Hochreiter, Jurgen Schmidhuberl. — 1997.
- [11] *Albert Gu Karan Goel, Christopher R'e.* Efficiently Modeling Long Sequences with Structured State Spaces / Christopher R'e Albert Gu, Karan Goel. — 2021.

- [12] *Albert Gu Tri Dao, Stefano Ermon Atri Rudra.* Hippo: Recurrent memory with optimal polynomial projections / Stefano Ermon Atri Rudra Albert Gu, Tri Dao, Christopher Ré. — 2020.
- [13] *Tsonis A.A. Deyle E.R., Ye H. Sugihara G.* Convergent Cross Mapping: Theory and an Example / Ye H. Sugihara G. Tsonis A.A., Deyle E.R. // Advances in Nonlinear Geosciences. — 2018.
- [14] *Roman Isachenko Ilya Zharikov, Artem Bochkarev Vadim Strijo.* Feature Generation for Physical Activity Classification / Artem Bochkarev Vadim Strijo Roman Isachenko, Ilya Zharikov. — 2018.