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

Физтех-школа прикладной математики и информатики
Кафедра «Интеллектуальные системы»
при Вычислительном центре им. А. А. Дородницына РАН)

Выпускная квалификационная работа бакалавра

Аппроксимация фазовой траектории временного ряда

Автор:
Студент 784 группы
Христолюбов Максим Евгеньевич

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

Москва 2022

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 универсальный алгоритм построение 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 про помощи мобильного акселерометра.

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

Оглавление

Глава 1

Введение

Одним из общих подходов к анализу временных рядов является взгляд на них как на измерение параметров детерминированные динамические 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. 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 литературе теневое многообразие называют фазовой траекторией ряда.

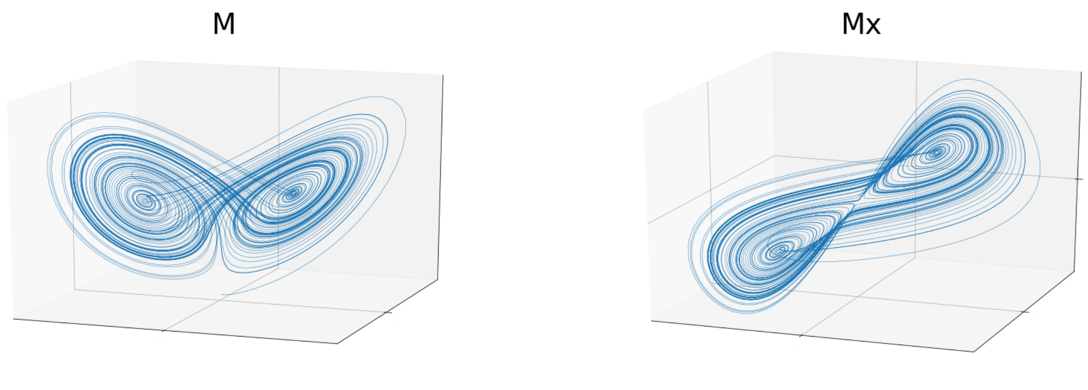

Рис. 1.1: Фазовая траектория системы $M = [X(t), Y(t), Z(t)]$ и теневая фазовая траектория $M_X = [X(t), X(t-\tau), X(t-2\tau)]$ [2]

Фазовая траектория временного ряда, точки которой являются предысториями каждого момента времени, имеет размерность N по построению. Распространенные способы отображении фазовой траектории в низкоразмерное пространство размерности n - PCA и PLS [3].

PLS проецирует матрицу фазовой траектории исходного временного ряда $\mathbf{X}^{(n)} \in \mathbb{R}^{m\times n}$, где m — количество моментов времени временного ряда, а n — размерность фазового пространства и целевую матрицу $\mathbf{Y}$ в скрытое пространство малой размерностью 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 $\mathbf{Y}$. При этом PLS максимизирует ковариацию между столбцами матриц T и U соответственно. Метод PLS соответствует следующей коммутативной диаграмме:

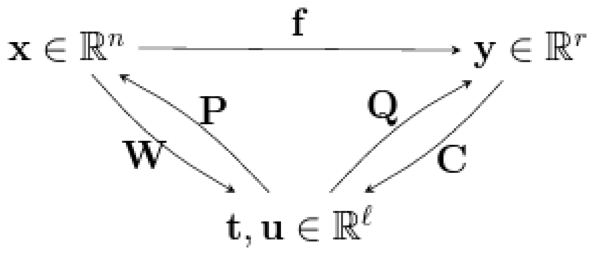

Рис. 1.2: диаграмма PLS

Кроме того, на большинство моделей временных рядов имеют интерпретацию с точки зрения аппроксимации фазовой траектории: авторегрессионная модель [4], сингулярное спектральное разложение (SSA) [5]. В качестве низкоразмерного пространства в них выступает пространство параметров модели временного ряда.

В данной работе исследуется модель S4, структура которой использует идею низкоразмерной проекции фазового пространства и динамически вычисляет ее без необходимости напрямую обращаться к предыстории каждого момента времени и построении фазового пространства фиксированной размерности n.

Глава 2

Постановка задачи

2.1 Фазовая траектория ряда

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 истинной 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.

2.2 Фазовая траектория ряда как гладкое многообразие

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

Определение 3. *[6] Гладким n-мерным многообразием* M *называется множество, для которого задана система подмножеств* X_i *и взаимно однозначные отображения на них* $\phi_i : W_i \to X_i$ *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* $W_{ij} = \phi_i^{-1}(X_i \cap X_j)$ *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* $W_{ij} = \phi_i^{-1}(X_i \cap X_j)$ *на* $W_{ji} = \phi_j^{-1}(X_i \cap X_j)$.

Определение 4. *Взаимнооднозначное отображение: $\phi : W \to X$, 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. Например, в роли динамической системы может выступать ходьба человека, а фазовая траектория строится по показаниям акселерометра на теле. Более подробно почему фазовая траектория — это многообразие показано в следующих разделах.

2.3 Алгоритм аппроксимации

Пусть $\mathbf{x} = [x_i]_{i=0}^m$ —- точка фазовой траектории, $f(t,\mathbf{w})$ — модель временного ряда.

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 $\boldsymbol{\xi} : \mathbb{X} \to \mathbb{W}$ объектов выборки $\mathbf{X} \in \mathbb{R}^N$ называется вложением $\mathbf{X}$ в скрытое пространство $\mathbb{W} = \mathbb{R}^n$*

Определение 8. *Аппроксимацией фазовой траектории называется композиция вложения точки фазовой траектории в скрытое пространство $\boldsymbol{\xi} : \mathbb{X} \to \mathbb{W}$ и восстановления $\phi : \mathbb{W} \to \mathbb{X}$.*

Общую структуру аппроксимации иллюстрирует диаграмма:

$$\begin{tikzcd}
\mathbb{X} \arrow[rr, bend left, "{[i]}"] \arrow[d, bend right, "\boldsymbol{\xi}"'] & \mathcal{T} \arrow[l, "\eta"'] \arrow[r, "s"'] \arrow[d] & \mathbb{S} \\
\mathbb{W} \arrow[u, bend right, "\phi"'] \arrow[r] & \mathcal{T} \times \mathbb{W} \arrow[ur, "f"'] &
\end{tikzcd}$$

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 $M = \cup X_i$. На практике нет доступа ко всему многообразию, а в наличии есть только конечный набор точек $\mathbf{X} = [\mathbf{x}_0, \dots, \mathbf{x}_m]^\intercal$.

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 ее окрестность и строится $\phi_i : W_i \to X_i$ так, что $\forall \mathbf{w} \in W_i : \phi_i(\mathbf{w}) = \hat{\mathbf{x}}_i$.

Таким образом из алгоритма аппроксимации естественным образом возникает система отображений $\phi_i : W_i \to X_i$ и обратных отображений $\phi_i^{-1} = \xi_i : X_i \to W_i$

2.4 Оптимальная размерность скрытого пространства

Размерность N фазового пространство, 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 $\mathbf{w} \in \mathbb{R}^n$. Возникает вопрос о том какая должна быть оптимальная размерность скрытого пространства и в каком смысле понимается оптимальность.

Поскольку скрытое пространство — 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 могу 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.

Утверждается, что оптимальная размерность скрытого пространства — это размерность многообразия, представляющего фазовую траекторию. Таким образом, если удалось построить гладко согласованный атлас, то размерность W_i является оптимальным размером скрытого пространства.

2.5 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_i 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 $\phi = [f(t_j, \mathbf{w})]_{j=o}^{N}$. Практически все существующие модели временных рядов $f(t, \mathbf{w})$ дифференцируемы 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 $\frac{\partial \mathbf{w}}{\partial \mathbf{x}} = \frac{\partial \boldsymbol{\xi}(\mathbf{x})}{\partial \mathbf{x}}$.

Для произвольной модели временного ряда $f(t,\mathbf{w})$ временного ряда вложение $\boldsymbol{\xi}$ строится, используя минимизацию функции ошибки. Эта техника используется в авторегрессионных моделях AR, ARIMA [7] [8], в фурье моделях и при аппроксимации сплайнами [9], SSA Гусеница.

$$\mathbf{w} = \boldsymbol{\xi}(\mathbf{x}) = \arg\min_{\mathbf{w}} L(\mathbf{x}, f(\mathbf{w}))$$

Для того, чтобы модель аппроксимации была корректной она должна быть гладко дифференцируемой, отсюда возникает вопрос о дифференцируемости $\boldsymbol{\xi}$.

Теорема 1. *Пусть дана функция*

$$\boldsymbol{\xi}(\mathbf{x}) = \arg\min_{\mathbf{w}} L(\mathbf{x}, f(\mathbf{w}))$$

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

Доказательство.

Согласно необходимому условию минимума функции L:

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

Дифференцируя равенство по $\mathbf{x}$, получим:

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

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

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

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

$$\partial^2_{\mathbf{w}} L(\mathbf{x}, f(\mathbf{w}))|_{\mathbf{w}=\boldsymbol{\xi}(\mathbf{x})} = f''(\mathbf{w}) \cdot \partial_f L(\mathbf{x}, f(\mathbf{w})) + (f'(\mathbf{w}))^2 \cdot \partial^2_f L(\mathbf{x}, f(\mathbf{w}))|_{\mathbf{w}=\boldsymbol{\xi}(\mathbf{x})}$$

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

Следовательно, при заданных условиях отображение $\boldsymbol{\xi}$ гладко дифференцируемо.

Таким образом, в модели ARIMA $\boldsymbol{\xi}$ 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 $\boldsymbol{\xi}$ вычисляется с помощью матричных операций и сингулярного разложения, что означает дифференцируемость $\boldsymbol{\xi}$.

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 точке нулевой меры, то проблем не будет. Множество корней уравнения $\partial^2_{\mathbf{w}} L(\mathbf{x}, f(\mathbf{w}))|_{\mathbf{w}=\boldsymbol{\xi}(\mathbf{x})} = 0$ является множеством меры 0, поэтому на практике построенное данным образом отображение не отличимо от гладкого отображения. Поэтому, согласованность атласа определяется исключительно непрерывностью $\xi \circ \phi$, из чего следует взаимнооднозначность ϕ.

Глава 3

Модели аппроксимаций

3.1 SSA Гусеница

Задача алгоритма SSA состоит в представлении сегмента временного ряда в виде суммы интерпретируемых компонент. Точке фазовой траектории $\mathbf{x}_i^N$ ставится в соответствие его траекторная матрица Ганкеля $\mathbf{X}_i^N \in R^{N/2 \times N/2+1}$:

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

где N — выбирается равным 2 характерным периодам ряда.

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

$$\mathbf{X}_i^{(N)} = \mathbf{U}(\mathbf{h})\mathbf{V} = \sum_{j=1}^{N/2} h_j \mathbf{u_j} \mathbf{v_j}^\mathsf{T}, \tag{3.1}$$

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 [10], 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 Автоенкодер LSTM

LSTM (long short-term memory, дословно (долгая краткосрочная память) — тип рекуррентной нейронной сети, способный обучаться долгосрочным зависимостям. LSTM были представлены в [11], 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.

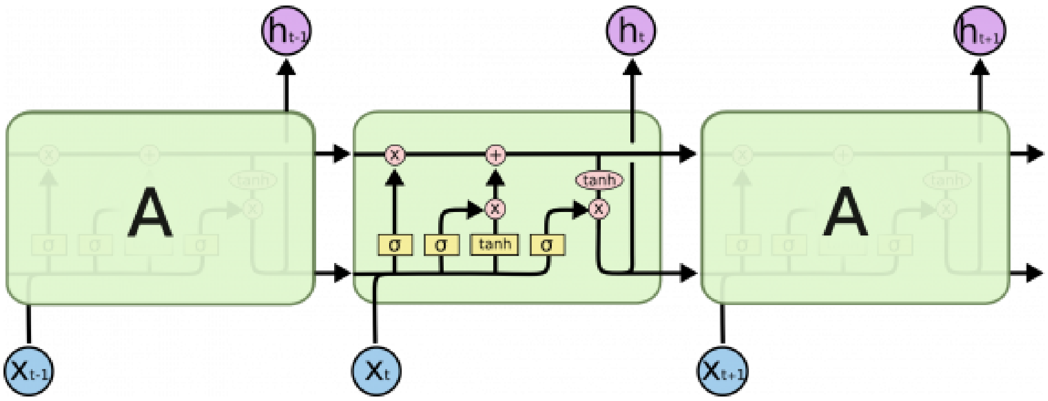

Рис. 3.1: Структура слоя LSTM

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

3.3 Neural ODE

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

$$\frac{dy}{dt} = g(y; w)$$

То есть, цель состоит не в том, чтобы изучить взаимосвязь между y и t, а в том, чтобы понять лежащую в основе динамику изменений. Если динамика не меняется слишком сильно, это обладает очень мощными возможностями обобщения. Прямой проход через нейронное ОДУ эквивалентен решению дифференциального уравнения с начальными условиями, в котором правая часть задается нейронной сетью $g(y; w)$. Это означает, что один прямой проход дает нам всю траекторию, в отличие, например, от RNNs, где каждый прямой проход через модель дает одно предсказание во времени. Решение ОДУ имеет вид:

$$y(t) = y(t_0) + \int_{t_0}^{t} g(y(t); w)dt$$

Нейросеть $g(y; w)$ 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. Информация о состоянии системы в данный момент содержится в предыстории, поэтому требуется, чтобы модель восстанавливала предысторию. Предлагается ввести на предыстории меру $\mu^{(\xi)}(t)$. Например, если $\mu^{(\xi)}(t) = \frac{1}{\xi}\mathbb{I}_{[t-\xi,t]}(t)$, тогда параметры модели в каждый момент времени t будет хранить информацию только о предыстории $[t-\xi,t]$. Данную меру называют translated Legendre measures (LegT). В статье [12] используется равномерный вес для всей истории, тогда мера $\mu^{(\xi)}(t) = \frac{1}{\xi}\mathbb{I}_{[0,\xi]}(t)$ (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 $||f||_{L2}(\mu) = \langle f, f\rangle_\mu^{\frac{1}{2}}$.

Тогда предыстория проецируется на n-размерное подпространство, при этом вложением фазовой траектории является разложение предыстории в ортогональном базисе.

Для скрытого состояния выполнено:

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

Для конкретного момента времени ξ, $x(t)$ находится с помощью скалярного произведения, но ее невозможно вычислять итеративно. Но функция коэффициента $\boldsymbol{\chi}(t)$ является решение ОДУ $\frac{d}{dt}\boldsymbol{\chi}(t) = \mathbf{A}(t)\boldsymbol{\chi}(t) + \mathbf{B}(t)x(t)$, 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 [12] в качестве базиса $\boldsymbol{\chi}(t)$, линейной оболочкой которого приближается предыстория, являются многочлены Лежандра $P_n(t)$. Известно, что они ортогональны относительно скалярного произведения, индуцированного мерой $\mathbf{1}_{[-1,1]}$:

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

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

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

При этом многочлены Лежандра удовлетворяют следующим соотношениям, используемых при вычислении коэффициентов базиса через скалярное произведение:

$$(2n+1)P_n(t) = P'_{n+1}(t) - P'_{n-1}(t), \quad P'_{n+1}(t) = (n+1)P_n(t) + tP'_n(t)$$

Алгоритм нахождения матрицы HIPPO (матрицы $\mathbf{A}$):

1) Продифференцировать равенство $\boldsymbol{\chi}_n(\xi) = \langle x, g^{(\xi)}\rangle_{\mu^{(\xi)}}$ по переменной ξ с помощью формулы Лейбница.

2) Привести полученное соотношение к виду $\frac{d}{d\xi}\boldsymbol{\chi}(\xi) = \mathbf{A}\boldsymbol{\chi}(\xi) + \mathbf{B}x(\xi)$

3) Дискретизировать ОДУ и получить рекуррентное соотношение.

3.4.2 Модель аппроксимации

Модель S4, используемая в [12], задается системой уравнений, в которой $x(t_0), y(t_0) \in \mathbb{R}, \boldsymbol{\chi}(t_0) \in \mathbb{R}^h$:

$$\boldsymbol{\chi}'(t) = \mathbf{A}\boldsymbol{\chi}(t) + \mathbf{B}x(t)$$

$$y(t) = \mathbf{C}\boldsymbol{\chi}(t) + \mathbf{D}x(t),$$

Матрица $\mathbf{A}$ иницилизируется 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 $\mathbf{A} = \mathbf{V\Lambda V}^* - \mathbf{PQ}^*$, где $\mathbf{V} \in \mathbb{C}^{h\times h}$, $\boldsymbol{\Lambda}$ диагональная и $\mathbb{P}, \mathbb{Q} \in \mathbb{R}^{h\times 1}$

Обучаемые параметры модели - диагонаяльная матрица $\boldsymbol{\Lambda}$ и векторы $\mathbf{P}$,$\mathbf{Q}$,$\mathbf{B}$,$\mathbf{C}$.

Дискретное представление получается из непрерывного с помощью дискретизация непрерывного времени по правилу $x_k = x(k\Delta t)$.

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

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

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

Пусть $\boldsymbol{\chi}_{-1} = 0$, тогда из реккурентного соотношения:

$\boldsymbol{\chi}_0 = \bar{\mathbf{B}}x_0,\ y_0 = \bar{\mathbf{C}}\bar{\mathbf{B}}x_0,$

$\boldsymbol{\chi}_1 = \bar{\mathbf{A}}\bar{\mathbf{B}}x_0 + \bar{\mathbf{B}}x_1,\ y_k = \bar{\mathbf{C}}\bar{\mathbf{A}}\bar{\mathbf{B}}x_0 + \bar{\mathbf{C}}\bar{\mathbf{B}}x_1$

$\boldsymbol{\chi}_1 = \bar{\mathbf{A}}^2\bar{\mathbf{B}}x_0 + \bar{\mathbf{A}}\bar{\mathbf{B}}x_1 + \bar{\mathbf{B}}x_2,\ y_k = \bar{\mathbf{C}}\bar{\mathbf{A}}^2\bar{\mathbf{B}}x_0 + \bar{\mathbf{C}}\bar{\mathbf{A}}\bar{\mathbf{B}}x_1 + \bar{\mathbf{C}}\bar{\mathbf{B}}x_2$

$y_k = \bar{\mathbf{C}}\bar{\mathbf{A}}^k\bar{\mathbf{B}}x_0 + \bar{\mathbf{C}}\bar{\mathbf{A}}^{k-1}\bar{\mathbf{B}}x_1 + \ldots + \bar{\mathbf{C}}\bar{\mathbf{B}}x_k$

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

Таким образом, $y_k = \bar{\mathbf{K}} * x$ — 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 $\bar{\mathbf{A}}$.

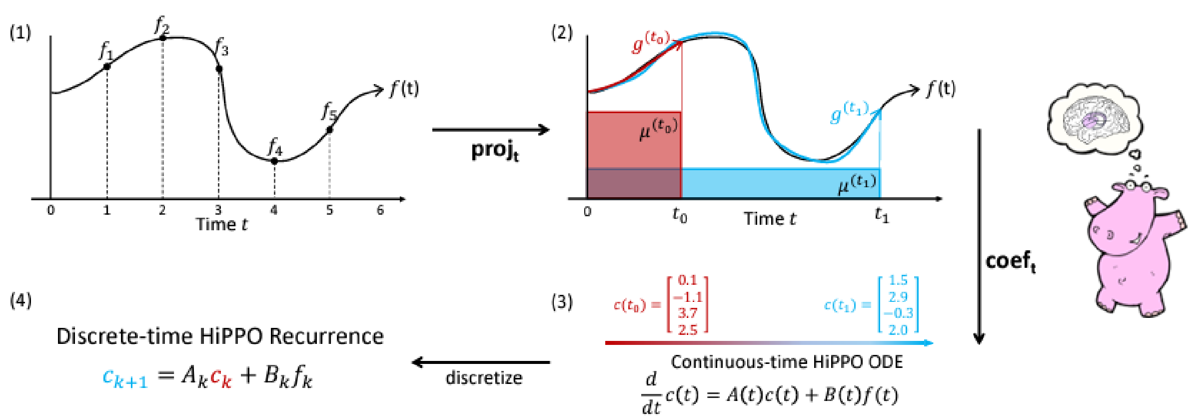

Рис. 3.2: диаграмма алгоритма HIPPO

Теорема 2. *Непрерывое и дискретное уравнение скрытого состояния системы для HIPPO-LegS:*

$$\frac{d}{dt}\boldsymbol{\chi}(t) = -\frac{1}{t}\mathbf{A}\boldsymbol{\chi}(t) + \frac{1}{t}\mathbf{B}x(t), \quad \boldsymbol{\chi}_{k+1} = \left(1 - \frac{\mathbf{A}}{k}\right)\boldsymbol{\chi}_k + \frac{1}{k}\mathbf{B}x_k$$

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

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

Глава 4

Описание практической части

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 Данные

Синтетический данные сгенерированы по формуле

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

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 датасетах объекты отсортированы по исследуемому признаку (a_i и вес рюкзака).

4.2 Критерии

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 - **число параметров**.

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 то 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:

a) Получить аппроксимацию $\hat{\mathbf{x}}(t)$ точки $\mathbf{x}(t)$.

b) Поскольку объекты выборки отсортированы по исследуемому признаку и близки друг к другу, то достаточно вычислить метрики

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

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

$$\mathrm{MaxStabError}(\mathbf{x}) = \max_i \left(\mathrm{StabError}(\mathbf{x}_i)\right)$$

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

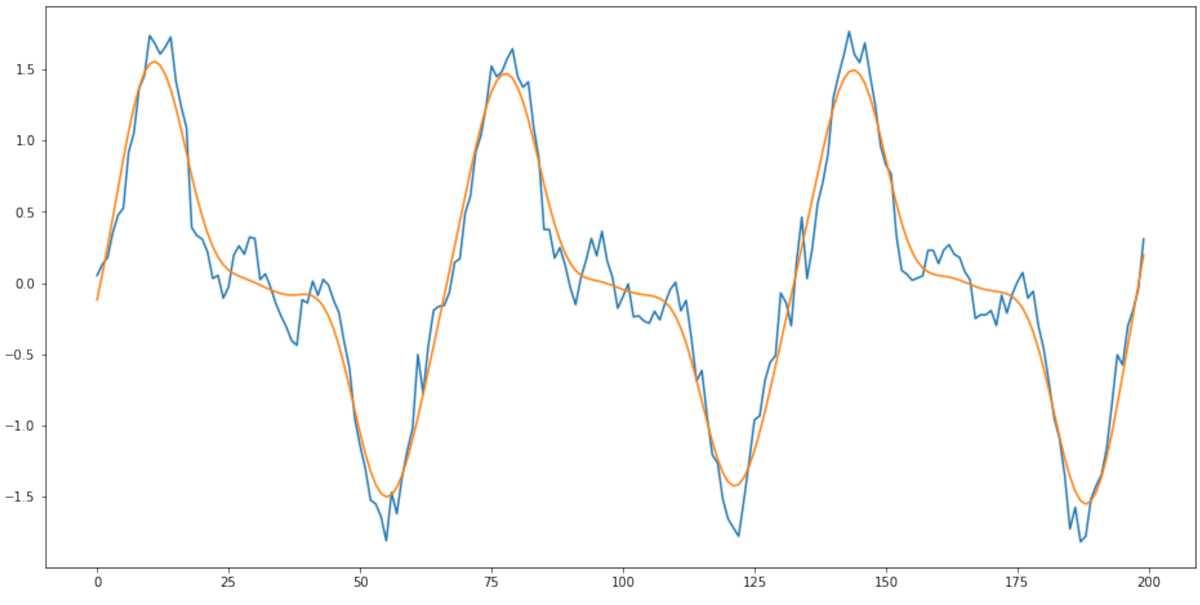

Рис. 4.1: Аппроксимация синтетического ряда

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

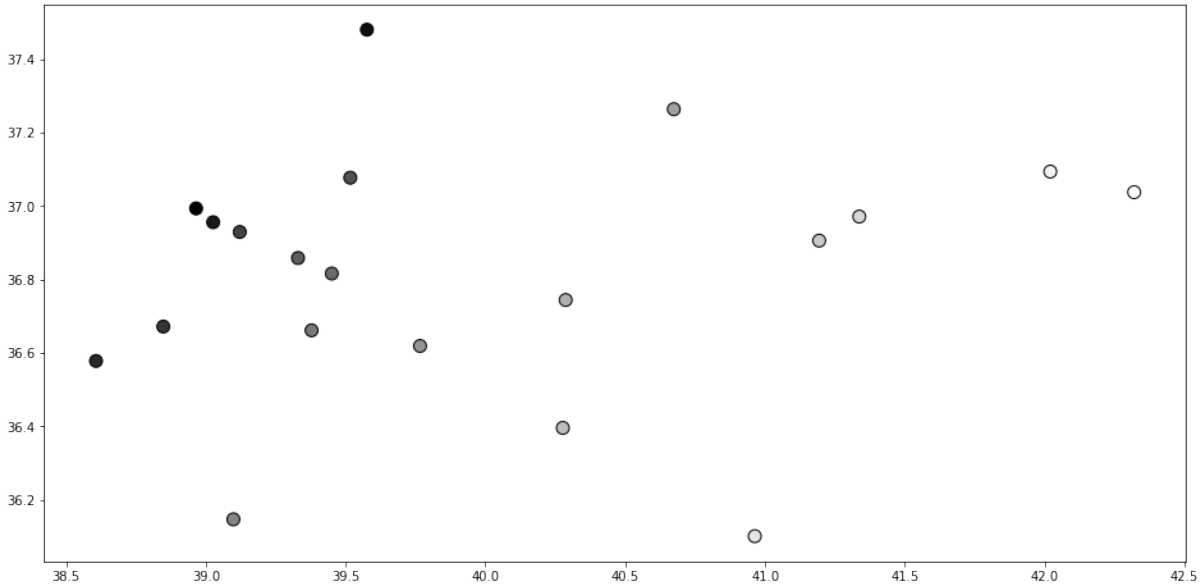

Рис. 4.2: Параметры SSA для разных рядов

Сравнение различных моделей аппроксимации:

Модель	Отклонение	Число параметров	MeanStabError	MaxStabError
SSA	1.77	2	2.12	3.67
LSTM				
S4				

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] *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.

[3] *В.В., Исаченко Р.В. Стрижов.* Снижение размерности пространства в задачах декодирования сигналов / Исаченко Р.В. Стрижов В.В. — 2021.

[4] *Lukashin, Y.P.* Adaptive methods of short-term forecasting of time series / Y.P. Lukashin // Finance and statisticss. — 2003.

[5] *Hassani, H.* Singular spectrum analysis: methodology and comparison. Journal of Data Science / H. Hassani. — 2007.

[6] *Чернавский, А.В.* / А.В. Чернавский // Часть первая. Многообразия. — 2010.

[7] *Agrawal, Ratnadip Adhikari R. K.* An Introductory Study on Time Series Modeling and Forecasting / Ratnadip Adhikari R. K. Agrawal. — 2013.

[8] *Н. В. Артамонов Е. А. Ивин, А. Н. Курбацкий Д. Фантаццини* / А. Н. Курбацкий Д. Фантаццини Н. В. Артамонов, Е. А. Ивин // ВВЕДЕНИЕ В АНАЛИЗ ВРЕМЕННЫХ РЯДОВ. — 2021.

[9] *Д. А. Аникеев Г. О. Пенкин, В. В. Стрижов.* Классификация физической актив ности человека с помощью локальных аппроксимирующих моделей / В. В. Стрижов Д. А. Аникеев, Г. О. Пенкин // Информ. и её применение. — 2019.

[10] *Andreas Kriegl Peter W. Michor, Armin Rainer.* Denjoy-Carleman differentiable perturbation of polynomials and unbounded operators / Armin Rainer Andreas Kriegl, Peter W. Michor. — 2009.

[11] *Sepp Hochreiter Jurgen Schmidhuberl, Peter W. Michor Armin Rainer.* LONG SHORTTERM MEMORY, Neural Computation / Peter W. Michor Armin Rainer Sepp Hochreiter, Jurgen Schmidhuberl. — 1997.

[12] *Albert Gu Karan Goel, Christopher R'e.* EFFICIENTLY MODELING LONG SEQUENCES WITH STRUCTURED STATE SPACES / Christopher R'e Albert Gu, Karan Goel. — 2021.

[13] *Albert Gu Karan Goel, Christopher R'e.* EFFICIENTLY MODELING LONG SEQUENCES WITH STRUCTURED STATE SPACES / Christopher R'e Albert Gu, Karan Goel // Вычислительные методы в физике плазмы / Ed. by Б. Олдера, С. Фернбаха, М. Ротенберга. — М.: Мир, 2021.

[14] *Roman Isachenko Ilya Zharikov, Artem Bochkarev Vadim Strijo*. Feature Generation for Physical Activity Classification / Artem Bochkarev Vadim Strijo Roman Isachenko, Ilya Zharikov. — 2018.