Федеральное государственное автономное образовательное учреждение высшего образования

«Московский физико-технический институт (национальный исследовательский университет)» Физтех-школа Прикладной Математики и Информатики Кафедра интеллектуальных систем

Направление подготовки / специальность: 03.03.01 Прикладные математика и физика (бакалавриат)

Направленность (**профиль**) **подготовки:** Компьютерные технологии и интеллектуальный анализ данных

ОПТИМАЛЬНОЕ ПОРОЖДЕНИЕ МОДЕЛЕЙ ЛОКАЛЬНОЙ АППРОКСИМАЦИИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ВРЕМЕННЫХ РЯДОВ

(бакалаврская работа)

Содержание

1	Вве	едение	4
2	Обзор литературы Постановка задачи классификации временных рядов		7
3			
4	Ло	кальная аппроксимация	9
	4.1	Модель авторегрессии $\mathrm{AR}(p)$	10
	4.2	Анализ сингулярного спектра	10
	4.3	Self-modeling regression	11
	4.4	Дискретное преобразование Фурье	12
5	Алгоритм обучения		13
	5.1	Построение признакового пространства Z	14
	5.2	Модель Цыбенко	15
	5.3	Выбор базовой модели	16
	5.4	Выбор моделей локальной аппроксимации	17
6	Данные		18
	6.1	Сегментация данных	18
	6.2	Нормализация сегментов	20
7	Вычислительный эксперимент		21
	7.1	Проверка статистических свойств полученной выборки	21
	7.2	Сложность модели	23
8	Вы	волы	24

Аннотация

Ключевые слова: классификация временных рядов; модели локальной аппроксимации; порождение признаков.

1 Введение

В работе исследуются методы построения пространства признаков временных рядов. При этом временной ряд рассматривается как последовательность сегментов, аппроксимируемых некоторой параметрической моделью, параметры которой используются в качестве их признаковых описаний[6]. Рассматривается задача анализа информативности признаков, полученных в результате применения различных аппроксимирующих моделей. Для порождения признаков используются модель авторегрессии, дискретное преобразование Фурье, модели ARMA, SSA, SEMOR [4,6].

Признаковая классификация временных рядов заключается в сопоставлении каждому сегменту временного ряда его признакового описания, позволяющему решать задачу классификации в пространстве признаков. Задача выбора оптимальной модели классификации, например, рассматривалась в[9].

Признаковое пространство будет порождаться по данным временным рядам с помощью аппроксимирующих моделей. Характерные временные интервалы разделим на 3 категории: элементарное движение (шаг), действие (ходьба), жизнь (прогулка). Решается задача классификация элементарных движений. Предлагается использовать сегментирование извлечение подпоследовательностей временного ряда, необходимые для описания элементарного движения максимальной продолжительности. Требуется построить суперпозицию моделей для построения пространства достаточной размерности, чтобы по нему можно было выполнить классификацию необходимого качества.

Способы сегментации периодических рядов описаны в [10]. Предлагаемый метод локально-аппроксимирующих моделей работает и для апериодических рядов. В [11] проведен обзор локально-аппроксимирующих

моделей. Предлагается применение смеси моделей и поиск наиболее подходящих для получения информации с использованием оптимальных параметров.

Для практической проверки предлагаемого метода в рамках вычислительного эксперимента решается задача классификации физической активности человека по измерениям нескольких датчиков. Эксперимент состоит из двух частей. В первой части осуществляется обработка временных рядов: порождение признаков с помощью локально-аппроксимирующих моделей. Во второй части происходит классификация на основе признаков, полученных суперпозицией исследуемых моделей. Оценивается сложность полученного классификатора. В итоге решена задача порождения пространства признакового описания с оптимальными свойствами и обеспечивающего классификацию временных рядов требуемого качества.

В данной работе мы рассматриваем принцип порождения локальноаппроксимирующих моделей, которые могут использоваться для порождения простой выборки, а именно случайной, однородной и независимой. Также предлагается метод оценивания *непростоты* модели с использованием критерия нелинейности.

2 Обзор литературы

В данной работе решается прямая задача, поставленную в [7]. Базовые алгоритмы обработки временных рядов в задаче классификации физической активности описаны в [5]. Использование описаний сегментов для порождения признакового пространства рассмотрено в работе [6]. Модель двухслойной нейронной сети в качестве базовой обсулавливается теоремой, доказанной в [12]

3 Постановка задачи классификации временных рядов

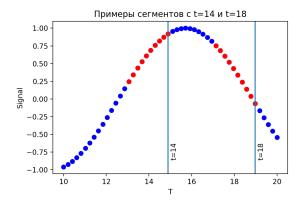
Пусть задан многомерный временной ряд – показания датчиков \mathbf{s} в моменты времени t из множества $\mathbb{T}=0,\ldots,T$. Можем говорить, что определена функция \mathbf{s} , отображающая дискретное пространство меток времени в непрерывное p -мерное пространство \mathbb{R}^{I} , где p – количество используемых датчиков.

$$\mathbf{s}: \mathbb{T} \to \mathbb{R}^{\scriptscriptstyle{|}}, |T| < \infty$$

Определим сегмент \mathbf{x}_t временного ряда по ширине сегмента T и метке времени t:

$$\mathbf{x}_t = vec(\mathbf{s}_t, \dots, \mathbf{s}_{t-T})$$

Рис. 1: Сегмент временного ряда с T = 10. (из работы [7])



Каждому моменту времени t ставится в соответсвие класс движения из множества \mathbb{Y} . Требуется найти отображение $f:t\to \hat{y}$, где $y\in\mathbb{Y}=1,2,\ldots,K$.

Предлагается в качестве исходного описания сегментов временных рядов использовать фазовое пространство $\mathbb{R}^{1 \times \mathbb{T}}$, порожденное главными

компонентами для каждого момента времени t.

В итоге решаем задачу классификации с обучающей выборкой $\mathfrak D$

$$\mathfrak{D} = \{(\mathbf{x}_i, y_i, t_i)\}_{i=1}^l,$$

где $\mathbf{x} \in \mathbb{R}^{p \times T}$ — точка фазового пространства сегментов данных с датчиков, $y_i \in \{1,2,\dots K\}$ — метки классов движения.

Выбирая в качестве функции потери multinomial log-loss:

$$L(f(x_i), y) = \sum_{i=1}^{l} \sum_{k=1}^{K} [y_i = k] \log P(y_i = k | x_i, \theta)$$
(3.1)

ставим оптимизационную задачу:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{l} L(f(x_i), y_i)$$

4 Локальная аппроксимация

Моделью локальной аппроксимации называется модель, аппроксимирующая временной ряд $\mathbf{x}(t)$ в промежутке времени [t,t-T]:

$$g:[t,t-T]\to \hat{\mathbf{x}}.$$

Или поэлементно $g: \tau \mapsto \mathbf{s}$. Метод локальной аппроксимации заключается в решении следующей задачи оптимизации. Оптимальные параметры $\hat{\mathbf{v}}$ метода локальной аппроксимации $g = g(\mathbf{v})$ определяются как $\hat{\mathbf{v}}(\mathbf{x}) = \arg\min_{\mathbf{v}} S(\mathbf{v}) = \|\mathbf{x} - \hat{\mathbf{x}}_g\|$, где $S(\mathbf{v})$ — функция потерь. Порожденные признаки \mathbf{z} — функция от решения задачи аппроксимации.

Таким образом, задан конечный набор моделей локальной аппроксимации **Ф**:

$$g \in \{g_1, \ldots, g_k\}$$
, где $g_j(\mathbf{v}, \mathbf{x}) \in \mathfrak{G} : [t, t - T] \mapsto \hat{\mathbf{x}}$

Термин «аппроксимирующая» означает, что модель приближает временной ряд \mathbf{x} в пространстве (\mathbb{X}, ρ) , то есть на некоторых оптимальных значениях $\mathbf{v} \in \mathbb{R}^n$

$$g(\mathbf{v}, \mathbf{x}) = \hat{\mathbf{x}}, \text{ где } \rho(\hat{\mathbf{x}}, \mathbf{x}) < \varepsilon.$$

При этом естественно взять в качестве признакового описания временного ряда \mathbf{x} вектор оптимальных параметров его модели.

Признаковым описанием временного ряда \mathbf{x} , порожденным параметрической моделью $g(\mathbf{v}, \mathbf{x})$ назовем вектор оптимальных параметров этой модели

$$\mathbf{v}_g(\mathbf{x}) = \operatorname*{arg\,min}_{\mathbf{v} \in \mathbb{R}^n} \rho\left(g(\mathbf{v}, \mathbf{x}), \mathbf{x}\right). \tag{4.1}$$

В качестве локально аппроксимирующих моделей предлагается использовать следующие модели.

4.1 Модель авторегрессии $\mathsf{AR}(p)$

Выберем в качестве модели локальной аппроксимации авторегрессионную модель порядка p:

$$g(\mathbf{w}, \mathbf{x}) = [\hat{x}^{(1)}, \dots, \hat{x}^{(t)}],$$
 где $\hat{x}^{(k)} = \begin{cases} x^{(k)}, & k = 1, \dots, p, \\ w_0 + \sum_{i=1}^p w_i x^{(k-i)}, & k = p+1, \dots, t. \end{cases}$ (4.2)

Тогда, выбрав в качестве ρ евклидово расстояние, получим признаковое описание ${\bf w}$ объекта ${\bf x}$:

$$\mathbf{w}_g(\mathbf{x}) = \underset{\mathbf{w} \in \mathbb{R}^n}{\min} \|\mathbf{x} - \hat{\mathbf{x}}_r\|_2^2 = \underset{\mathbf{w} \in \mathbb{R}^n}{\arg\min} \|\mathbf{x} - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1} \mathbf{X}^\mathsf{T}\mathbf{x}.$$
(4.3)

4.2 Анализ сингулярного спектра

В качестве альтернативной гипотезы порождения данных рассмотрим модель ряда SSA [8]. Поставим в соответствие временному ряду $\mathbf{x} = \{x^{(1)},...,x^{(t)}\}$ траекторную матрицу \mathbf{X} , то есть, матрицу следующего вида:

$$\mathbf{X} = \begin{pmatrix} x^{(1)} & x^{(2)} & \dots & x^{(n)} \\ x^{(2)} & x^{(3)} & \dots & x^{(n+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x^{(m-n+1)} & x^{(m-n+2)} & \dots & x^{(m)} \end{pmatrix}. \tag{4.4}$$

Построим сингулярное разложение матрицы $\mathbf{X} \uparrow \mathbf{X},$

$$\mathbf{X} \uparrow \mathbf{X} = \mathbf{VHV} \uparrow$$
, $\mathbf{H} = \operatorname{diag}(h_1, ..., h_n)$,

где $h_1, ..., h_n$ — собственные числа матрицы $\mathbf{X} \uparrow \mathbf{X}$, отвечающие за величины различных частота спектра временного ряда \mathbf{x} . Для временного ряда \mathbf{x} рассматривается вектор сингулярных чисел \mathbf{h} в качестве нового признакового описания и строиться классификатор \hat{y}_i в виде

$$\hat{y}_i = a(\boldsymbol{\theta}, \mathbf{h}),$$

где оптимальные параметры $\hat{m{ heta}}$ минимизируют ошибку оптимизационной задачи при решении задачи классификации.

Для того, чтобы описать метод ручного выделения признаков, необходима постановка прикладной задачи и экспертные знания о структуре объектов. Прикладную задачу классификации данных акселерометра и метод экспертного выделения признаков рассмотрим следующем разделе.

4.3 Self-modeling regression

Будем считать, что кроме сегмента временного ряда x существует почти совпадающий с ним временной ряд p с меньшим количеством элементов. Модель SEMOR работает в предположении, что форма временного ряда p повторяет форму сегмента временного ряда x. Исходя из этого предположения можно аппроксимировать временной ряд x сегментом p.

Модель SEMOR задается как функция от \mathbf{x}, \mathbf{w} :

$$g(x, w) = w_1 + w_2 p(w_3 + w_4 t)$$

В общем случае p — функция формы, однако в частном случае это может быть заранее заданный временной ряд. Параметры w_1, w_2 находятся шагом двухпараметрической линейной регрессии, w_3, w_4 получаются минимизацией DTW, где из наклона пути, результирующего в minimum edit distance мы получаем w_4 , а w_3 — это смещение этой прямой относительно нуля. В итоге имеем оптимальные веса модели:

$$w_{\text{SEMOR}} = [\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \rho]$$

4.4 Дискретное преобразование Фурье

. Задан временной ряд

$$x = [x^{(0)}, \dots, x^{(t-1)}], \ x^{(k)} \in \mathbb{C}, \ k = 0, \dots, t - 1.$$

Взяв в качестве аппроксимирующей модели обратное преобразование Фурье,

$$g(\mathbf{v}, x) = [\hat{x}^{(0)}, \dots, \hat{x}^{(t-1)}],$$

где

$$\hat{x}^{(k)} = \frac{1}{t} \sum_{j=0}^{t-1} (v_{2j} + iv_{2j+1}) e^{\frac{2\pi i}{t}kj}, \ k = 0, \dots, t-1,$$
 (4.5)

получим, что признаковым описанием временного ряда x является прямое преобразование:

$$\mathbf{v}_g(x) = [v_0, \dots, v_{2t-1}], \text{ где } v_{2k} + iv_{2k+1} = \sum_{j=0}^{t-1} x^{(j)} e^{-\frac{2\pi i}{t}kj}, \ k = 0, \dots, t-1.$$

$$(4.6)$$

Переписывая 4.5 в матричном виде, заметим, что как и в предыдущих случаях, параметры модели **v** эквивалентно находятся при помощи линейной регрессии временного ряда на столбцы матрицы Фурье. Выбор лишь некоторых комплексных амплитуд соответствует регрессии временного ряда на соответствующие столбцы матрицы Фурье.

Заметим, что в первых двух случаях используются билинейные аппроксимирующие модели $g(\mathbf{v},x)$, а в третьем — линейная. Приведенные примеры демонстрируют большую общность построения пространства признаков при помощи моделей типа (??) и решения оптимизационной задачи (4.1). Вообще говоря, при $|X| \geqslant 2$ любая процедура построения признаковых описаний $\mathbf{g}: X \to \mathbb{R}^n$ задается эквивалентно решением оптимизационной задачи (4.1) при выборе соответствующей пары (g, ρ) .

5 Алгоритм обучения

В качестве исходного признакового пространства, как уже было сказано ранее, используется фазовое пространство исходных сегментов временных рядов. Нормализованные сегменты и будем считать точками данного фазового пространства.

Подавая на вход базовой модели точки данного фазового пространства будем получать сразу метку класса движения, характерного для момента времени с данной фазовой траекторией. Получим базовую точность, на которую будем ориентироваться при обучении $k_{\mathbf{h}}$ моделей локальной аппроксимации \mathbf{g} , мы применяем для многоклассовой классификации двухслойную нейронную сеть полной связи с K выходными слоями.

Алгоритм 5.1 Алгоритм обучения

```
Вход: \mathbf{X}, \mathbf{Y}, \mathbf{g}, k_{\mathbf{g}};
```

Выход: $\hat{\theta}$;

1: инициализировать $\mathbf{u} := \mathbf{y}_1$ (вектор матрицы \mathbf{Y})

2:
$$\mathbf{z} = []$$

з: для $i=1,\dots,k_{\mathbf{g}}$

$$\mathbf{z}[i] := \mathbf{g}[i](\mathbf{X})$$

4: инициализировать $\mathbf{Z} = \mathbf{z}[1]$

5: для
$$i=2,\ldots,k_{\bf g}$$

$$\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$$

6: $\hat{\theta} = \arg\min(L(f(Z), Y))$

Алгоритм 5.2 Применение алгоритма

Вход: $\mathbf{X}, \mathbf{h}, k, k_{\mathbf{g}}, \hat{\theta};$

Выход: $\hat{\mathbf{y}}$;

1: инициализировать $\mathbf{u} := \mathbf{y}_1$ (вектор матрицы \mathbf{Y})

- 2: $\mathbf{z} = []$
- 3: для $i=1,\dots,k_{\mathbf{g}}$
- 4: $\mathbf{z}[i] := \mathbf{h}[i](\mathbf{X})$
- 5: инициализировать $\mathbf{Z} = \mathbf{z}[1]$
- 6: для $i=2,\ldots,k_{\mathbf{g}}$
- 7: $\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$
- 8: инициализировать $\hat{\mathbf{y}}$
- 9: для $i = 1, \dots, k$
- 10: $\hat{\mathbf{y}}[i] = \arg\max_{k \in KP(f(\mathbf{X}[i]) = y_k|\hat{\theta})}$

5.1 Построение признакового пространства Z

Итак, набор g состоит из локально аппроксимирующих моделей, каждая из которых отображает сегмент x в оптимальные параметры модели. Имеем следующее отображение

$$[\mathbf{g}_1 \dots \mathbf{g}_k] : x \mapsto [\mathbf{v}_1^* \dots \mathbf{v}_k^*]$$

Взятые в совокупности эти отображения порождают промежуточное признаковое пространство Z, элементы которого являются конкатенацией векторов $\mathbf{g}_1(x), \mathbf{g}_2(x) \dots \mathbf{g}_k(x)$.

Функция потерь определена на отображении a, то есть

$$L(f(\mathbf{x}), y) = L(a(g(\mathbf{x}), \gamma), y)$$

Итак, задача решается следующим образом:

ullet Поиск и вычисление отображения $Z = \{ \mathbf{g}(x_i) \}_{i=1}^l$ путем минимиза-

ции функционалов ошибок для каждой модели локальной аппроксимации и поиска оптимальных параметров

$$\arg\min_{\mathbf{w}\in W} L_g(\mathbf{X}, \mathbf{w}) = \arg\min_{\mathbf{w}\in W} \sum_{i=1}^l \sum_{k=1}^n ||g(\mathbf{w}, \mathbf{x}_i) - \mathbf{x}_i||_2^2$$

• Оптимизация функции ошибки обобщенной линейной модели, которая в качестве обучающей выборки имеет (\mathbf{Z}, y) .

$$\arg\min_{\theta\in\Theta} L_a(\mathbf{Z}, \mathbf{y}, \theta) = \arg\min_{\theta\in\Theta} \left[-\sum_{i=1}^l \sum_{k=1}^K [y_i = k] \log P(y_i = k | \mathbf{z}_i, \theta) \right]$$

Основное допущение, принимаемое в данном методе является допущение о том, что выборка в признаковом пространстве объектов является простой. В данной работе мы рассматриваем, для каких признаковых пространств это допущение справедливо, а также предлагаем способы построения таких выборок.

5.2 Модель Цыбенко

Цель работы - оптимальный отбор моделей локальной аппроксимации, порождающих простую выборку. Мы хотим получить оптимальные признаки, порожденные выбранными моделями локальной аппрокскимации, для простой модели классификации. Поэтому выбрано решение исследовать сложность, порождаемую исследуемыми моделями. Для классификации предлагается использовать двухслойную нейронную сеть, последний слой которой представляет обобщенную линейную модель, а количество нейронов на скрытом слое будет обозначать сложность выборки.

Th. Цыбенко, 1989

Пусть $\sigma(\xi)$ любая непрерывная сигмоидная функция, например, $\sigma(\xi) = 1/(1+e^{-\xi})$. Тогда, если дана любая непрерывная функция действительных переменных f на $I_n = [0,1]^n$ и $\epsilon > 0$, тогда существуют векторы \mathbf{w} и параметризована функция $G(\mathbf{x}, \sigma, \mathbf{w}) : I_n \to \mathbb{Y}$ такая, что $|G(\mathbf{x}, \mathbf{w}) - f(\mathbf{x})| < |\epsilon|$ для всех $\mathbf{x} \in I_n$, где $\mathbf{w}_i \in \mathbb{R}^n$, $G = \sum_i \sigma(\mathbf{w}_i^T \mathbf{x})$.

5.1. Критерием нелинейности будем называть сложность модели, а именно количество нейронов на скрытом слое двухслойной нейронной сети (модели "Цыбенко"), а обозначать будем $Comp(\mu)$, где μ - обучаемая модель:

$$Comp(\mu) = \#|neurons\ in\ the\ hidden\ layer|$$

5.3 Выбор базовой модели

Каждая модель классификации характеризуется тремя показателями: точностью, устойчивостью и сложностью. В данной работе выбор моделей локальной аппроксимации обсулавливается минимизацией сложности модели, порождающей простую выборку. Поэтому в алгоритме обучения модели "Цыбенко"будет использоваться фиксированная точность и устойчивость, определямые базовой моделью. В качестве базовой модели предлагается использовать многослойную нейронную сеть. На вход данной сети подаются объекты из фазового пространства $\mathbb{R}^{I \times T}$. Выходом являются вероятности меток классов движения, по которым определяется наиболее вероятный класс для фазовой траектории метки временного ряда.

Было исследовано несколько архитектур нейронных сетей с разным количеством скрытых слоев. Для (N-1)— слойной нейронной сети исследовалось оптимальное количество нейронов на последнем скрытом слое. Затем архитектура фиксировалась и обучалась N— слойная нейронная

сеть, где варьировалось количество на предпоследнем слое. В итоге выбрана оптимальная по сложности модель двухслойной нейронной сети, на точность S^* которой будет ориентироваться при обучении модель "Цыбенко".

5.4 Выбор моделей локальной аппроксимации

- 1. Наращиваем сложность модели "Цыбенко"до достижения заданной точности аппроксимации S*, полученной с помощью базовой модели.
- 2. Выбираем лучшую по сложности модель.
- 5.2. Определение структуры модели "Цыбенко

Пусть $\mathcal{A} \in \mathcal{J}$ — структура модели "Цыбенко": $\mathcal{A} = \{\mathbf{z}_i\}$, где $i \in \mathcal{I} \subset \mathcal{J}, \mathcal{J}$ — подмножество индексов выбранных на текущей итерации моделей локальной аппроксимации.

5.3. Оптимальная структура

$$\mathcal{A}^* = \arg\min_{\mathcal{A} \in \mathcal{J}} Comp(\mathcal{A})$$

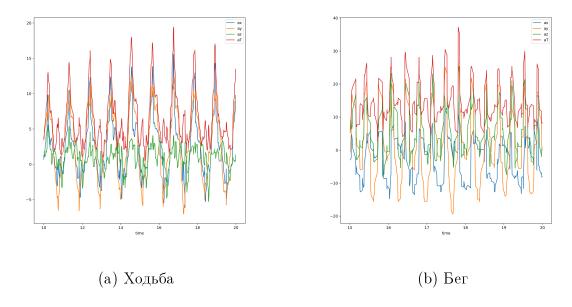


Рис. 2: Профиль данных 1. (из работы [7])

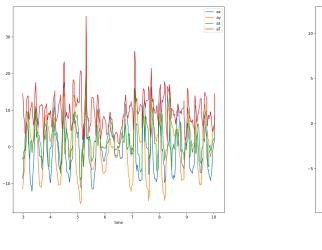
6 Данные

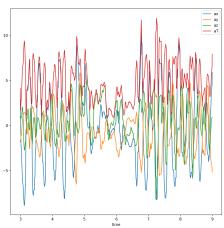
В работе используется датасет, собранный автором работы [7] Исходные данные представляют собой набор временных рядов, соответствующих 4 классам: бег, ходьба, подъем и спуск по лестнице. Исходные временные ряды сегментируются и нормализуются, а затем отображаются в фазовое пространство.

6.1 Сегментация данных

Существует несколько методов для сегментации временных рядов. Метод "каблука" ориентируется на пики временного ряда и требует подготовки при сборе данных. В работе [10] рассматривается автоматический подход для сегментации данных.

В данной работе используется основанный на предположении о средней постоянной частоте действий подход. По заранее предполагаемой средней длительности действия F и его временному ряду S(t) ищутся локальные максимумы соответсвующего радиуса F.





(а) Вверх по лестнице

(b) Вниз по лестнице

Рис. 3: Профиль данных 2.(из работы [7])

$$\mathbf{I} = \{i : S(i) = \max_{j=i-F...i+F} \{S(j)\}\}\$$

Полученные индексы будут использованы для инициализации начала текущего сегмента и конца предыдущего.

Алгоритм 6.1 Сегментация временного ряда действия

Вход: S, l, F;

Выход: х;

- 1: I = []
- 2: k = 0
- 3: для $i \in 1 \dots l$
- 4: если $S(i) = \max_{j=i-F...i+F} \{S(j)\}$ то
- 5: $\mathbf{I} = (\mathbf{I}, i)$
- 6: k = k + 1
- 7: $\mathbf{x} = []$
- 8: для $i \in 1 \dots k-1$
- 9: $\mathbf{x} = (\mathbf{x}, \mathbf{S}[\mathbf{I}[k] : \mathbf{I}[k+1]])$

6.2 Нормализация сегментов

Фазовые траектории полученных сегментов могут отличаться для различных действий по количеству главных информативных фазовых компонент. Для нормализации используются кубические сплайны. Подробная процедура нормализации описана в [7].

В итоге получена обучающая выборка $(\hat{\mathbf{x}}, Y)$, на которой проводятся вычислительные эксперименты.

7 Вычислительный эксперимент

Для проведения эксперимента были выбраны следующие локальноаппроксимирующие модели:

- SSA с параметром 5;
- AR-авторегрессия с параметром 4;
- FFT с 5 старшими частотами;
- SEMOR ходьба;
- SEMOR бег;
- SEMOR ходьба вверх по лестнице;
- SEMOR ходьба вниз по лестнице.

У каждой из вышеперечисленных моделей было по 5 структурных параметров, использующихся для порождения промежуточного признакового пространства. Перед каждой итерацией генетического алгоритма было по 10 'особей'-структур \mathcal{A} модели 'Цыбенко'.

7.1 Проверка статистических свойств полученной выборки

В зависимости от способа порождения данных, одна и та же выборка может быть простой, а может и не быть. Мы проверим некоторые гипотезы, выполняющиеся для простых выборок. То есть в случае невыполнения какой-либо гипотезы, мы с большой вероятносью можем утверждать, что выборка не является простой. Выполнение же гипотез не

будет противоречить нашему предположению о простоте выборки. Проверим свойства из определения простой выборки: независимость, случайность и однородность распределения. В нашей работе мы проверяли следующие гипотезы, также рассматриваемые в [7]:

• Равенство матожидания нулю и постоянство дисперсия ошибки, где ошибка есть

$$\varepsilon_i = \hat{y_i} - y_i$$

Проверка гипотезы осуществляется путем вычислением ошибки и использованием t-критерия Стьюдента.

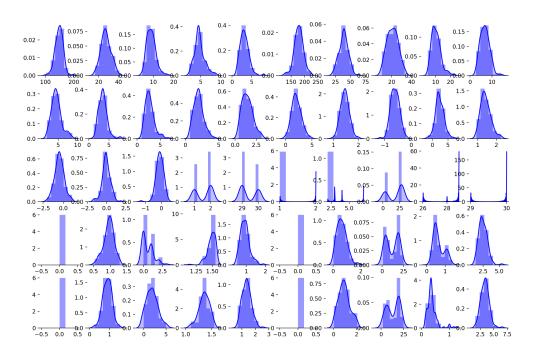
• Унимодальность распределения признаков. Проверим распределения всех признаков, выбранных в результате работы алгоритма выбора оптимальных моделей локальной аппроксимации.

На графиках видно, что большинство признаков имеют унимодальное распределение. Может не отвергать предположение о простоте выборки.

7.2 Сложность модели

Предположение, что промежуточное пространство обладает свойствами простой выборки, определяет наш выбор обобщенной линейной модели GLM в качестве последнего слоя модели "Цыбенко"с малым количеством нейронов на скрытом слое. GLM должна оказаться простой и интерпретируемой моделью по сравнению с базовой моделью, применяемой к исходному фазовому пространству и имеющей большое число нейронов. Достигнута высокая точность при малых ресурсных затратах. Таким образом, показана эквивалентность использования компози-

Рис. 4: Распределения признаков.



ции нескольких простых локально-аппроксимирующих моделей и полносвязной нейронной сети.

8 Выводы

Предложен критерий нелинейности выборки и метод выбора локальноаппроксимирующих моделей, порождающих простую выборку. Проведены эксперименты для построения универсальной модели, названной моделью "Цыбенко".

Для исследуемых классов временных рядов выбраны оптимальные модели локальной аппроксимации, а также оптимальная структура универсальной модели. Подтверждена гипотеза об избыточности промежуточного пространства, порожденного большим количеством локально-аппроксимирующих моделей.

Планируется расширить библиотеку локальных моделей аппроксимации для улучшения многообразия "популяции"структур модели 'Цыбенко'. Также плаируется сравнить используемый в работе подход порождения структуры модели 'Цыбенко', основанный на генетическом алгоритме, с иными методами многомерной оптимизации.

- 1 Argyro Kampouraki, George Manis, and Christophoros Nikou. Heartbeat time series classication with support vector machines. In IEEE Transactions on Information Technology in Biomedicine, volume 13, pages 512–518, 2009.
- 2 Damian Eads, Daniel Hill, Sean Davis, Simon Perkins, Junshui Ma, Reid Porter, and James Theiler. Genetic algorithms and support vector machines for time series classication. Proceedings of SPIE The International Society for Optical Engineering, 4787:74–85, 2002.
- 3 Michael Husken and Peter Stagge. Recurrent neural networks for time series classication, 2003.
- 4 Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao. Time series classication using multi-channels deep convolutional neural networks. In Lecture Notes in Computer Science (including subseries Lecture Notes in Articial Intelligence and Lecture Notes in Bioinformatics), volume 8485 LNCS, pages 298–310, 2014.
- 5 Кузнецов МП and Ивкин НП. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию. Машинное обучение и анализ данных, pages 1471–1483, 2015.
- 6 Карасиков ME and Стрижов BB. Классификация временных рядов в пространстве параметров порождающих моделей. Информ. и её примен., раде 121–131, 2016.
- 7 Иванычев МЕ. Выбор моделей локальной аппроксимации для классификации временных рядов. Выпускная квалификационная работа бакалавра, 2018.
- 8 H. Hassani. Singular Spectrum Analysis: Methodology and Comparison. Journal of Data Science, 5:239–257, 2007.

- 9 Задаянчук АИ, Попова МС and Стрижов ВВ, Выбор оптимальной модели классификации физической активности по измерениям акселерометра, Информационные технологии, 22(4), 313-318, 2016
- 10 Motrenko AP and Strijov VV, Extracting fundamental periods to segment human motion time series, Journal of Biomedical and Health Informatics, 2016
- 11 Лоскутов АЮ and Котляров ОЛ and Истомин ИА and Журавлев ДИ, Проблемы нелинейной динамики 111. Локальные методы прогнозирования временных рядов, Вестн. Моск. ун-та. Сер. 3. Физ. Астрон., 2002
- 12 Cybenko GV, Approximation by Superpositions of a Sigmoidal function, Mathematics of Control Signals and Systems, vol. 2 no. 4 pp. 303-314, 1989