

Классификация временных рядов в пространстве параметров порождающих моделей*

Карасиков М. Е.

karasikov@phystech.edu

Московский физико-технический институт

Работа посвящена задаче многоклассовой признаковой классификации временных рядов. Признаковая классификация временных рядов заключается в сопоставлении каждому временному ряду его краткого признакового описания, позволяющего решать задачу классификации в пространстве признаков. 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. Построенное так признаковое описание сегмента временного ряда наследует от модели аппроксимации такие полезные свойства, как инвариантность относительно сдвига. Для решения задачи классификации в качестве признаковых описаний временных рядов предлагается использовать распределения параметров аппроксимирующих сегменты моделей, что обобщает базовые методы, использующие непосредственно сами параметры аппроксимирующих моделей. Проведен ряд вычислительных экспериментов на реальных данных, показавших высокое качество решения задачи многоклассовой классификации. Эксперименты показали превосходство предлагаемого метода над базовым и многими распространенными методами классификации временных рядов на всех рассмотренных наборах данных.

Ключевые слова: *временные ряды; многоклассовая классификация; сегментация временных рядов; гиперпараметры аппроксимирующей модели; модель авторегрессии; дискретное преобразование Фурье.*

Feature-Based Time Series Classification*

Karasikov M. E.

Moscow Institute of Physics and Technology

1 Введение

Временным рядом x будем называть конечную упорядоченную последовательность чисел:

$$x = [x^{(1)}, \dots, x^{(t)}].$$

Временные ряды являются объектом исследования в таких задачах анализа данных, как прогнозирование [1], обнаружение аномалий [2], сегментация [3], кластеризация [4] и классификация [3]. Обзор по задачам и методам анализа временных рядов дается в [5, 6]. Последние годы связаны с ростом интереса к данной области, проявляющимся в непрекращающемся предложении новых методов анализа временных рядов — метрик алгоритмов сегментации кластеризации и других.

В данной работе рассматривается задача классификации временных рядов, возникающая во многих приложениях (медицинская диагностика по ЭКГ [7] и ЭЭГ [8, 9], класси-

фикация типов физической активности по данным с акселерометра [10, 11], верификация динамических подписей [12] и т. д.).

Формально задача классификации в общем виде ставится следующим образом. Пусть X — множество описаний объектов произвольной природы, Y — конечное множество меток классов. Предполагается существование целевой функции — отображения $y : X \rightarrow Y$, значения которого известны только на объектах обучающей выборки

$$\mathfrak{D} = \{(x_1, y_1), \dots, (x_m, y_m)\} \subset X \times Y.$$

Требуется построить классификатор $a : X \rightarrow Y$ — отображение, приближающее целевую функцию y на множестве X . При $|Y| > 2$ задачу классификации будем называть многоклассовой. Задачей классификации временных рядов будем называть задачу классификации, в которой объектами классификации являются временные ряды.

Задание метрики или функции расстояния [13] на парах временных рядов позволяет применять метрические методы классификации. При удачном выборе метрики классификация может производиться простейшими метрическими алгоритмами классификации, например, методом ближайшего соседа [14]. Данный подход к решению задачи классификации временных рядов чрезвычайно распространен в силу того, что позволяет свести исходную задачу классификации временных рядов к задаче выбора метрики.

Второй подход к решению задачи классификации состоит в построении для каждого временного ряда его информативного признакового описания $\mathbf{f} : X \rightarrow \mathbb{R}^n$, позволяющего строить точные классификаторы с хорошей обобщающей способностью. Построение информативного пространства признаков исходных объектов множества X , позволяющего добиться заданной точности классификации и значительно упрощающий последующий анализ, является важнейшим этапом решения задачи классификации. Признаки могут задаваться экспертом. Так в работе [15] предлагается использовать в качестве признаков статистические функции (среднее, отклонения от среднего, коэффициенты эксцесса и др.). Стоит заметить, что при таком подходе к построению пространства признаков часто удается добиться необходимого качества классификации путем выбора соответствующих конкретной задаче признаков (см. пример [16]), а сам выбор признаков становится важной технической задачей. Другой метод построения пространства признаков заключается в задании параметрической регрессионной или аппроксимирующей модели временного ряда. Тогда в качестве признаков временных рядов будут выступать параметры настроенной модели. В работе [17] в качестве признаков предлагается использовать коэффициенты дискретного преобразования Фурье (DFT), в [17, 18] — дискретного вейвлет-преобразования (DWT), а в [19, 20, 21] — модели авторегрессии. Таким образом, при данном методе построения признаков возникает задача выбора аппроксимирующей модели временного ряда.

В работе исследуются методы классификации временных рядов, использующие в качестве их признаков описаний параметры аппроксимирующих моделей. Приводится сравнение моделей аппроксимации. Из временного ряда могут извлекаться сегменты — его подпоследовательности, для которых признаковые описания, строятся так же, как и для исходных временных рядов. Использование подпоследовательностей позволяет обобщить алгоритмы классификации. Так в работе [3] предлагается алгоритм классификации временных рядов методом голосования их случайных сегментов (непрерывных подпоследовательностей со случайным начальным элементом). В нашей работе предлагается алгоритм классификации временных рядов в пространстве параметров распределений признаков

их сегментов, который сравнивается с родственным ему алгоритмом голосования сегментов [3]. В разд. 7 (Вычислительный эксперимент) приводятся эксперименты на реальных данных, показывающие высокое качество и общность предлагаемого алгоритма в сочетании с методом признаковых описаний временных рядов параметрами аппроксимирующих их моделей.

2 Постановка задачи

Поставим задачу многоклассовой классификации временных рядов в общем виде. Пусть (X, ρ) — метрическое пространство временных рядов, Y — множество меток классов, $\mathfrak{D} \subset X \times Y$ — конечная обучающая выборка.

Пусть

1. S — процедура сегментации:

$$S(x) \subset 2^{\mathbf{S}(x)}, \quad (1)$$

где $\mathbf{S}(x)$ — множество всех сегментов временного ряда $x \in X$,

2. \mathbf{f} — процедура построения признакового описания набора сегментов:

$$\mathbf{f}(S(x)) \in \mathbb{R}^n, \quad (2)$$

3. b — алгоритм многоклассовой классификации:

$$b : \mathbb{R}^n \rightarrow Y. \quad (3)$$

Рассмотрим семейство $A = \{a : X \rightarrow Y\}$ алгоритмов классификации вида

$$a = b \circ \mathbf{f} \circ S. \quad (4)$$

Пусть задана функция потерь

$$\mathcal{L} : X \times Y \times Y \rightarrow \mathbb{R}$$

и функционал качества

$$Q(a, \mathfrak{D}) = \frac{1}{|\mathfrak{D}|} \sum_{(x,y) \in \mathfrak{D}} \mathcal{L}(x, a(x), y). \quad (5)$$

В качестве методов обучения $\mu(\mathfrak{D}) \in A$ будем использовать следующие:

$$\mu_{\mathbf{f}, S}(\mathfrak{D}) = \hat{b} \circ \mathbf{f} \circ S,$$

где \hat{b} — минимизатор эмпирического риска:

$$\hat{b} = \arg \min_b Q(b \circ \mathbf{f} \circ S, \mathfrak{D}).$$

Оптимальный метод обучения определяется по скользящему контролю:

$$\mu^* = \arg \min_{\mathbf{f}, S} \widehat{CV}(\mu_{\mathbf{f}, S}, \mathfrak{D}),$$

где $\widehat{CV}(\mu, \mathfrak{D})$ — внешний критерий качества метода обучения μ ; при этом исходная обучающая выборка \mathfrak{D} случайно разбивается r раз на обучающую и контрольную ($\mathfrak{D} = \mathfrak{L}_1 \sqcup \mathfrak{T}_1 = \dots = \mathfrak{L}_r \sqcup \mathfrak{T}_r$),

$$\widehat{CV}(\mu, \mathfrak{D}) = \frac{1}{r} \sum_{v=1}^r Q(\mu(\mathfrak{L}_v), \mathfrak{T}_v), \quad (6)$$

где

$$Q(a, \mathfrak{T}) = \frac{1}{|\mathfrak{T}|} \sum_{(x,y) \in \mathfrak{T}} \mathbf{1}\{a(x) = y\}. \quad (7)$$

Средняя точность (precision) классификации объектов класса $c \in Y$ оценивается функционалом скользящего контроля (6) с модифицированным функционалом качества Q :

$$Q_c(a, \mathfrak{T}) = \frac{|\{(x, y) \in \mathfrak{T} \mid a(x) = y = c\}|}{|\{(x, y) \in \mathfrak{T} \mid y = c\}|}. \quad (8)$$

3 Сегментация временных рядов

Определение 1. Сегментом временного ряда $x = [x^{(1)}, \dots, x^{(t)}]$ будем называть любую его непрерывную подпоследовательность $s = [x^{(i)}]_{i=t_0}^{t_1}$, $1 \leq t_0 \leq t_1 \leq t$.

Определение 2. Под сегментацией будем понимать отображение временных рядов во множество их сегментов (выр. 1).

Примеры.

- Тривиальная сегментация

$$S(x) = \{x\}, \quad \forall x \in X. \quad (9)$$

- Случайное выделение сегментов некоторой длины ℓ [3].
- Важным является случай квазипериодичности временного ряда, когда сам ряд состоит из похожих в определенном смысле сегментов, называемых периодами:

$$x = \left[\underbrace{x^{(1)}, \dots, x^{(t_1)}}_{s^{(1)}}, \underbrace{x^{(t_1+1)}, \dots, x^{(t_2)}}_{s^{(2)}}, \dots, \underbrace{x^{(t_{p-1}+1)}, \dots, x^{(t)}}_{s^{(p)}} \right]. \quad (10)$$

Тогда в качестве процедуры сегментации можно взять разбиение на периоды:

$$S(x) = \{s^{(1)}, \dots, s^{(p)}\}. \quad (11)$$

Для выделения периодов могут быть использованы, например, алгоритмы [22, 23, 24].

4 Аппроксимирующая модель сегмента временного ряда

Поскольку сегмент временного ряда сам является временным рядом, в этом разделе слово сегмент будем опускать.

Определение 3. Параметрической аппроксимирующей моделью временного ряда x будем называть отображение

$$g : \mathbb{R}^n \times X \rightarrow X. \quad (12)$$

В слово «аппроксимирующая» вкладывается тот смысл, что модель должна приближать временной ряд в пространстве (X, ρ) , то есть для некоторого $\mathbf{w} \in \mathbb{R}^n$

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

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

Определение 4. Признаковым описанием временного ряда x , порожденным параметрической моделью $g(\mathbf{w}, x)$ назовем вектор оптимальных параметров этой модели

$$\mathbf{w}_g(x) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \rho(g(\mathbf{w}, x), x). \quad (13)$$

В качестве аппроксимирующих моделей предлагается использовать следующие.

- **Модель линейной регрессии.** Пусть задан r -компонентный временной ряд (например, время и 3 пространственные координаты):

$$x = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)}], \text{ где } \mathbf{x}^{(k)} = [x_1^{(k)}, \dots, x_r^{(k)}]^\top, \quad k = 1, \dots, t.$$

Рассмотрим модель линейной регрессии одной из компонент временного ряда на остальные компоненты как аппроксимирующую модель:

$$g(\mathbf{w}, x) = [\hat{\mathbf{x}}^{(1)}, \dots, \hat{\mathbf{x}}^{(t)}], \text{ где } \hat{\mathbf{x}}^{(k)} = [x_1^{(k)}, \dots, x_{r-1}^{(k)}, \hat{x}_r^{(k)}]^\top, \quad k = 1, \dots, t,$$

$$\underbrace{\begin{bmatrix} \hat{x}_r^{(1)} \\ \vdots \\ \hat{x}_r^{(t)} \end{bmatrix}}_{\hat{\mathbf{x}}_r} = \underbrace{\begin{bmatrix} x_1^{(1)} & \dots & x_{r-1}^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(t)} & \dots & x_{r-1}^{(t)} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_{r-1} \end{bmatrix}}_{\mathbf{w}}.$$

Тогда, выбрав в качестве ρ евклидово расстояние, по определению 4 получим признаковое описание объекта x :

$$\mathbf{w}_g(x) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{x}_r - \hat{\mathbf{x}}_r\|_2^2 = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{x}_r - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{x}_r. \quad (14)$$

- **Модель авторегрессии AR(p).** Задан временной ряд

$$x = [x^{(1)}, \dots, x^{(t)}], \quad x^{(k)} \in \mathbb{R}, \quad k = 1, \dots, t.$$

Выберем в качестве модели аппроксимации авторегрессионную модель порядка p :

$$g(\mathbf{w}, x) = [\hat{x}^{(1)}, \dots, \hat{x}^{(t)}], \text{ где } \hat{x}^{(k)} = \begin{cases} x^{(k)}, & k = 1, \dots, p, \\ w_0 + \sum_{i=1}^p w_i x^{(k-i)}, & k = p+1, \dots, t. \end{cases} \quad (15)$$

Далее признаковое описание определяется аналогично случаю линейной регрессии (см. [выр. 14](#)).

— **Дискретное преобразование Фурье.** Задан временной ряд

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

Взяв в качестве аппроксимирующей модели обратное преобразование Фурье,

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

где

$$\hat{x}^{(k)} = \frac{1}{t} \sum_{j=0}^{t-1} (w_{2j} + iw_{2j+1}) e^{\frac{2\pi i}{t} kj}, \quad k = 0, \dots, t-1, \quad (16)$$

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

$$\mathbf{w}_g(x) = [w_0, \dots, w_{2t-1}], \quad \text{где } w_{2k} + iw_{2k+1} = \sum_{j=0}^{t-1} x^{(j)} e^{-\frac{2\pi i}{t} kj}, \quad k = 0, \dots, t-1. \quad (17)$$

Переписывая [выр. 16](#) в матричном виде, заметим, что как и в предыдущих случаях, параметры модели \mathbf{w} 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. Случай дискретного вейвлет-преобразования аналогичен.

Заметим, что в первых двух случаях используются билинейные аппроксимирующие модели $g(\mathbf{w}, x)$, а в третьем — линейная. Приведенные примеры демонстрируют большую общность построения пространства признаков при помощи моделей типа [\(12\)](#) и решения оптимизационной задачи [\(13\)](#). Вообще говоря, при $|X| \geq 2$ любая процедура построения признаков описаний $\mathbf{f} : X \rightarrow \mathbb{R}^n$ задается эквивалентно решением оптимизационной задачи [\(13\)](#) при выборе соответствующей пары (g, ρ) .

5 Распределения признаков сегментов

Объединим идеи, изложенные в [разд. 3 \(Сегментация временных рядов\)](#) и [разд. 4 \(Аппроксимирующая модель сегмента временного ряда\)](#). Согласно аппроксимирующей модели [\(12\)](#) получим для каждого сегмента $s^{(k)} \in S(x) = \{s^{(1)}, \dots, s^{(p)}\}$ временного ряда x его признаковое описание $\mathbf{w}^{(k)} := \mathbf{w}_g(s^{(k)})$, решив оптимизационную задачу [\(13\)](#). Тогда всему набору сегментов $S(x)$ соответствует выборка

$$\mathbf{F} = (\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(p)}). \quad (18)$$

Примем гипотезу простоты выборки [\(18\)](#).

Гипотеза 1. Выборка $\mathbf{F} = (\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)})$ — простая, то есть случайная, независимая и однородная, где $\mathbf{w}^{(k)} \sim \mathbf{P}_0$.

Пусть имеется параметрическое семейство распределений $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$. Будем рассматривать вероятностную модель, в которой объект x зависит от случайного параметра θ .

Гипотеза 2. $p(x|\theta, y) = p(x|\theta)$.

Тогда

$$p(x, y) = p(\mathbf{F}, y) = \int_{\Theta} p(\mathbf{F}, \boldsymbol{\theta}, y) d\boldsymbol{\theta} = \int_{\Theta} p(\mathbf{F}|\boldsymbol{\theta}) p(\boldsymbol{\theta}, y) d\boldsymbol{\theta}.$$

При этом распределение $p(\boldsymbol{\theta}, y)$ предлагается оценивать на этапе обучения, где признаковыми описаниями объектов x_i задачи классификации являются оценки параметров $\boldsymbol{\theta}_i$:

$$\hat{\boldsymbol{\theta}}_i = T(x_i) = T(\mathbf{F}_i).$$

Получив оценку $\hat{p}(\boldsymbol{\theta}, y)$, находим оценку плотности $\hat{p}(x, y)$:

$$\hat{p}(x, y) = \int_{\Theta} p(\mathbf{F}|\boldsymbol{\theta}) \hat{p}(\boldsymbol{\theta}, y) d\boldsymbol{\theta},$$

по которой строится байесовский классификатор.

В алгоритмической постановке задачи классификации получим $\hat{p}(y|\boldsymbol{\theta}) = \delta(a(\boldsymbol{\theta}), y)$ и

$$\hat{p}(\boldsymbol{\theta}, y) = \delta(a(\boldsymbol{\theta}), y) p(\boldsymbol{\theta}).$$

Тогда

$$\begin{aligned} \hat{p}(x, y) &= \int_{\Theta} p(\mathbf{F}|\boldsymbol{\theta}) \hat{p}(\boldsymbol{\theta}, y) d\boldsymbol{\theta} = \\ &= \int_{\Theta} p(\mathbf{F}|\boldsymbol{\theta}) \delta(a(\boldsymbol{\theta}), y) p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{a^{-1}(y)} p(\mathbf{F}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \\ &= \int_{a^{-1}(y)} p(\boldsymbol{\theta}|\mathbf{F}) p(\mathbf{F}) d\boldsymbol{\theta}. \end{aligned}$$

Приближая распределение $p(\boldsymbol{\theta}|\mathbf{F})$ вырожденным $\delta(\boldsymbol{\theta} - T(\mathbf{F}))$, получим

$$\hat{p}(y|x) = \int_{a^{-1}(y)} p(\boldsymbol{\theta}|\mathbf{F}) d\boldsymbol{\theta} = \int_{a^{-1}(y)} \delta(\boldsymbol{\theta} - T(\mathbf{F})) d\boldsymbol{\theta} = \delta(a(T(\mathbf{F})), y).$$

Таким образом, задача классификации временных рядов свелась к задаче классификации оценок параметров распределений семейства $\{\mathbf{P}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \Theta}$.

В качестве оценок параметров $\boldsymbol{\theta}$ предлагается брать оценки максимального правдоподобия

$$\hat{\boldsymbol{\theta}} = T(x) = \arg \max_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta} | x) = \arg \max_{\boldsymbol{\theta} \in \Theta} p(\mathbf{F}|\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \prod_k p(\mathbf{w}^{(k)} | \boldsymbol{\theta}).$$

Заметим, что в частном случае тривиальной сегментации (9) и семейства вырожденных распределений оценка $\hat{\boldsymbol{\theta}}$ является исходным признаковым описанием. Таким образом, предложенный подход к построению признакового описания временного ряда

$$\mathbf{f} : x \mapsto \hat{\boldsymbol{\theta}} \tag{19}$$

является достаточно общим и при этом хорошо интерпретируется.

6 Алгоритм классификации

Для завершения построения классификатора временных рядов (4) построим многоклассовый классификатор b (см. [выр. 3](#)) по обучающей выборке $\{(\mathbf{f}(x), y) \mid (x, y) \in \mathfrak{D}\}$.

Сведем задачу многоклассовой классификации к задачам бинарной классификации. Наиболее общей стратегией к сведению является Error-Correcting Output Codes [25], которая обобщает стратегии One-vs-All и One-vs-One.

В нашей работе для решения задач бинарной классификации, где $Y = \{-1, +1\}$, берутся регуляризованная логистическая регрессия (RLR) и различные модификации SVM.

— Классификатор **SVM** выглядит следующим образом:

$$f(x; \mathbf{w}, w_0) = \text{sign}(\mathbf{w}^\top \mathbf{f}(x) - w_0),$$

где параметры \mathbf{w} и w_0 определяются решением задачи безусловной минимизации

$$\frac{1}{2C} \|\mathbf{w}\|_2^2 + \sum_{(x,y) \in \mathfrak{D}} \max\{1 - y(\mathbf{w}^\top \mathbf{f}(x) - w_0), 0\} \rightarrow \min_{\mathbf{w} \in \mathbb{R}^n, w_0 \in \mathbb{R}}.$$

— Линейный классификатор **RLR** записывается в виде

$$f(x; \mathbf{w}) = \text{sign} \mathbf{w}^\top \mathbf{f}(x),$$

где вектор параметров \mathbf{w} определяется из условия

$$\frac{1}{2C} \|\mathbf{w}\|_2^2 + \sum_{(x,y) \in \mathfrak{D}} \log(1 + e^{-y\mathbf{w}^\top \mathbf{f}(x)}) \rightarrow \min_{\mathbf{w} \in \mathbb{R}^n}.$$

7 Вычислительный эксперимент

Вычислительный эксперимент проводился на данных для задачи классификации типов физической активности человека.

7.1 Датасет WISDM

Датасет WISDM [11] содержит показания акселерометра для шести видов человеческой активности:

- | | |
|-------------|---------------|
| 1. Jogging | 4. Downstairs |
| 2. Walking | 5. Sitting |
| 3. Upstairs | 6. Standing |

Необработанные данные, представляющие из себя последовательность размеченных показаний акселерометра (по тройке чисел на каждый отсчет времени с интервалом в 50 миллисекунд), были разбиты на временные ряды длиной по 200 отсчетов (10 секунд).

Классы	Jogging	Walking	Upstairs	Downstairs	Sitting	Standing
Число объектов	1624	2087	549	438	276	231

Таблица 1. Распределение временных рядов по классам. Dataset: WISDM.

7.1.1 Ручное выделение признаков

Выбор признаков. В первом эксперименте в качестве признаков описаний временных рядов использовались их статистические функции. Каждая компонента временного ряда описывалась 40 чисел — её средним, стандартным отклонением, средним модулем отклонения от среднего, гистограммой с 10 областями равной ширины. Полученные признаки для каждой компоненты объединялись и к ним добавлялся признак средней величины ускорения.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One. В качестве бинарного классификатора использовался SVM с RBF ядром и параметрами $C = 8.5$, $\gamma = 0.12$.

Результаты. На диаграмме ниже (рис. 1) приведено качество классификации, усредненное по $r = 50$ случайным разбиениям исходной выборки на тестовую и контрольную в пропорциях 7 к 3.

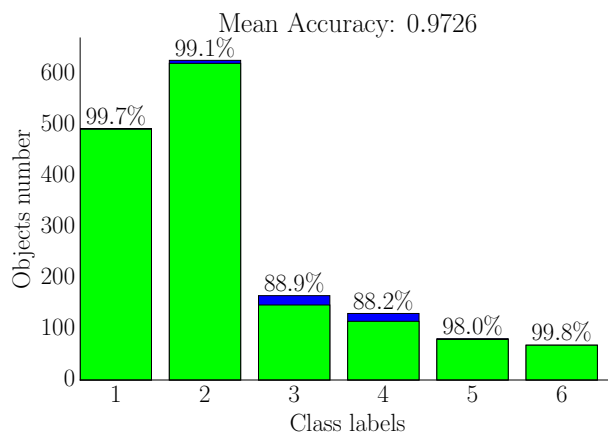


Рис. 1. Точность классификации при ручном выделении признаков. Dataset: WISDM. Под Mean ассигасу понимается значение функционала (7). Над столбцами приведены средние точности классификации для каждого класса по формуле (8).

Как видно из таблицы 2, классы 2, 3 и 4 недостаточно хорошо отделяются друг от друга.

		Predicted class					
		1	2	3	4	5	6
Actual class	1	1.00	0.00	0.00	0.00	0.00	0.00
	2	0.00	0.99	0.01	0.00	0.00	0.00
	3	0.03	0.04	0.89	0.04	0.00	0.00
	4	0.02	0.05	0.05	0.88	0.00	0.00
	5	0.01	0.00	0.00	0.00	0.98	0.00
	6	0.00	0.00	0.00	0.00	0.00	1.00

Таблица 2. Mean confusion matrix. Ручное выделение признаков. Dataset: WISDM.

7.1.2 Модель авторегрессии (15)

Признаковое описание. Во втором эксперименте в качестве признаков описаний временных рядов использовались все статистические функции, что брались в первом эксперименте, разд. 7.1.1 (Выбор признаков), за исключением гистограммы. Вместо 10 значений для каждого блока гистограммы использовались 7 коэффициентов модели авторегрессии AR(6) (см. выпр. 15). Таким образом, каждый временной ряд описывался 31 числами. Также проводилась предварительная нормализация признаков.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-All. В качестве бинарного классификатора использовался SVM с RBF ядром и параметрами $C = 8$, $\gamma = 0.8$.

Результаты. На диаграмме ниже приведено качество классификации, усредненное по $r = 50$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

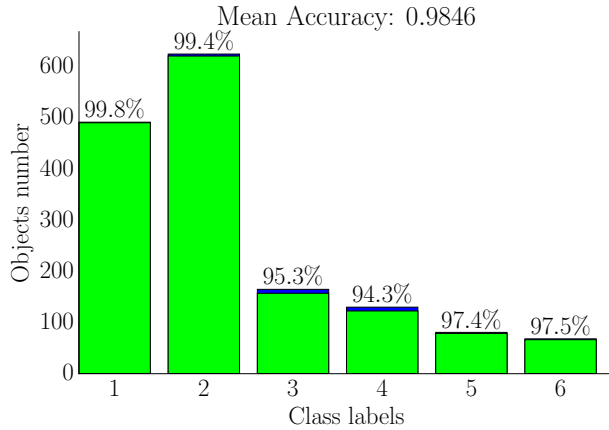


Рис. 2. Точность классификации для параметров модели авторегрессии в качестве признаков описаний. Dataset: WISDM. Под Mean ассигасу понимается значение функционала (7). Над столбцами приведены средние точности классификации для каждого класса по формуле (8).

Несмотря на неравномерное распределение объектов по классам, использование признакового описания, порожденного моделью авторегрессии, позволяет значительно повысить качество классификации. Точность построенного классификатора минимальна для 4 класса «Downstairs» и составляет 94.3%. Алгоритм классификации можно улучшать добавлением дополнительных признаков, например, параметров линейной регрессии, однако, добиваться максимально возможной точности не входит в цели нашего эксперимента.

		Predicted class					
		1	2	3	4	5	6
Actual class	1	1.00	0.00	0.00	0.00	0.00	0.00
	2	0.00	0.99	0.00	0.00	0.00	0.00
	3	0.01	0.02	0.95	0.02	0.00	0.00
	4	0.00	0.02	0.04	0.94	0.00	0.00
	5	0.01	0.00	0.00	0.00	0.97	0.01
	6	0.01	0.00	0.00	0.00	0.01	0.97

Таблица 3. Mean confusion matrix. Признаки, порожденные моделью авторегрессии. Dataset: WISDM.

7.2 Датасет USC-HAD

Датасет USC-HAD [26] содержит показания акселерометра для 12 типов физической активности человека:

1. walk forward
2. walk left
3. walk right
4. go upstairs
5. go downstairs
6. run forward
7. jump up and down
8. sit and fidget
9. stand
10. sleep
11. elevator up
12. elevator down

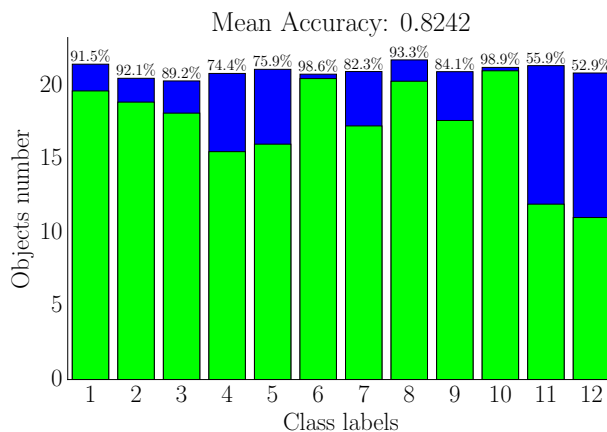
Выборка содержит примерно по 70 шести-компонентных временных ряда для каждого класса, а средняя длина временного ряда порядка 3300. Частота записи измерений сенсора 100 Hz.

7.2.1 Ручное выделение признаков

Выбор признаков. В качестве признаков брались те же признаки, что и в предыдущем эксперименте, [разд. 7.1.1 \(Выбор признаков\)](#).

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One. В качестве бинарного классификатора использовался SVM с RBF ядром и параметрами $C = 80$, $\gamma = 0.002$.

Результаты. Исходная выборка 100 раз случайно разбивалась на обучающую и контрольную в отношении 7 к 3. На диаграмме [3](#) приведен результат — процент верной классификации для объектов каждого класса.



		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	0.92	0.00	0.00	0.03	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	2	0.01	0.92	0.01	0.01	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.01	0.89	0.06	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00
	4	0.03	0.02	0.11	0.74	0.07	0.01	0.02	0.00	0.00	0.00	0.00	0.00
	5	0.05	0.08	0.00	0.06	0.76	0.00	0.05	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00
	7	0.03	0.00	0.00	0.02	0.05	0.08	0.82	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.02	0.01	0.02	0.02
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.84	0.00	0.08	0.04
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99	0.00	0.00
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.56	0.40
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.41	0.53

Рис. 3. Точность классификации при ручном выделении признаков. Dataset: USC-HAD. Под Mean ассурасу понимается значение функционала (7). Над столбцами приведены средние точности классификации для каждого класса по формуле (8).

Из таблицы [4](#) видно, что классы 11 и 12 («elevator up» и «elevator down») плохо отделяются друг от друга, то есть статистические признаки не достаточно чувствительны, чтобы разделить эти классы. Также ошибка на классах 4 и 5 («go upstairs» и «go downstairs») превышает 20%.

Таблица 4. Mean confusion matrix. Ручное выделение признаков. Dataset: USC-HAD.

7.2.2 Модель авторегрессии (15)

Признаковое описание. При записи данных USC-HAD сенсор делал каждую секунду 100 измерений. Предполагая, что на каждое «элементарное движение» человек тратит порядка секунды, приходим к выводу, что параметры авторегрессионной модели малых порядков в данном случае неинформативны. Приведем исходные временные ряды к частоте 10 Hz при помощи осреднения.

В качестве признаков описаний преобразованных временных рядов брались статистические функции, описанные в [разд. 7.1.1 \(Выбор признаков\)](#), за исключением гистограммы. Так же для каждой компоненты отдельно и для модуля результирующего ускорения и поворота добавлялось по 11 параметров авторегрессионной модели AR(10) (см. [выр. 15](#)). Затем проводилась нормализация признаков.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-All. В качестве бинарного классификатора использовался SVM с RBF ядром и параметрами $C = 16$, $\gamma = 0.1$.

Результаты. На диаграмме [4](#) ниже приведено качество классификации, усредненное по $r = 200$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

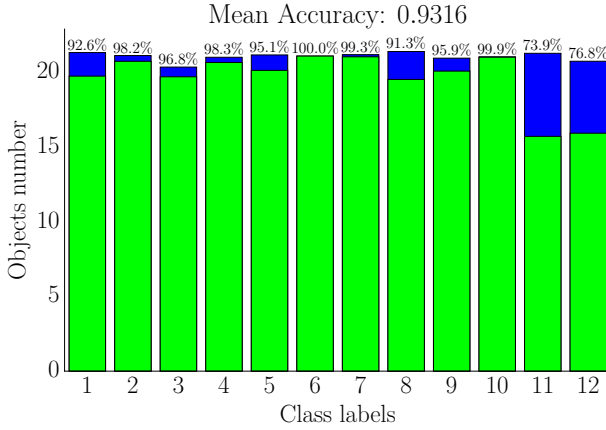


Рис. 4. Точность классификации для параметров модели авторегрессии в качестве признаков описаний. Dataset: USC-HAD. Под Mean accuracy понимается значение функционала (7). Над столбцами приведены средние точности классификации для каждого класса по формуле (8).

Из таблицы [5](#) видно, что использование признаков описания, порожденного моделью авторегрессии, значительно повысило качество классификации для всех классов. Недостаточно достоверно отделяются только классы 11 и 12 («elevator up» и «elevator down»), где ошибка составляет 25%. Однако, они превосходно отделяются от всех остальных классов.

		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	0.93	0.02	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.02	0.01	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.03	0.00	0.00	0.00	0.95	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.91	0.07	0.00	0.02	0.00	0.00
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.96	0.00	0.02	0.00	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.26	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.21	0.77	0.23

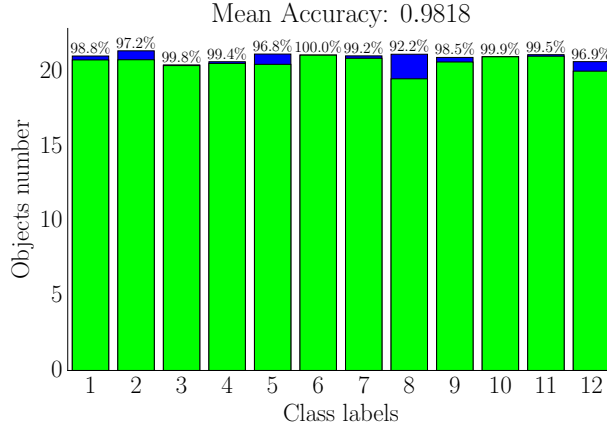
Таблица 5. Mean confusion matrix. Признаки, порожденные моделью авторегрессии. Dataset: USC-HAD.

7.2.3 Модель авторегрессии (15) и Фурье (17)

Признаковое описание. Признаковые описания временных рядов брались из предыдущего эксперимента, [разд. 7.2.2 \(Признаковое описание\)](#), с добавлением к ним 3 — 12 коэффициентов Фурье (17). Таким образом, каждый 6-компонентный временной ряд описывался 128 признаками.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One. В качестве бинарного классификатора использовался SVM с RBF ядром и параметрами $C = 10$, $\gamma = 0.13$.

Результаты. На диаграмме 5 ниже приводится качество классификации, усредненное по $r = 500$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.



		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	0.99	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.01	0.97	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.01	0.97	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.08	0.00	0.00	0.00
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.97

Рис. 5. Точность классификации для параметров модели авторегрессии в качестве признаков описаний. Dataset: USC-HAD. Под Mean ассигасу понимается значение функционала (7). Над столбцами приведены средние точности классификации для каждого класса по формуле (8).

Из таблицы 6 видно, что использование коэффициентов Фурье значительно повысило качество классификации. Хуже всего класс 8 («sit and fidget») отделяется от класса 9 («stand»). Точность классификации для него составляет 92.2%.

Таблица 6. Mean confusion matrix. При- знаки, порожденные моделью авторегрессии. Dataset: USC-HAD.

7.2.4 Классификация голосованием и классификация в пространстве распределений параметров

Рассмотрим, алгоритм классификации в сочетании с процедурой сегментации временных рядов. В качестве процедуры сегментации $S(x)$ (см. **выр. 1**) будем использовать выделение сегментов фиксированной длины. Решим задачу классификации для первых 10 классов (за исключением «elevator up» и «elevator down», которые плохо отделяются друг от друга при малой длине сегментов) двумя алгоритмами.

В алгоритме голосования классификатор $b : \mathbb{R}^n \rightarrow Y$ обучается на новой обучающей выборке для сегментов исходных временных рядов

$$\mathfrak{D}_S = \{(\mathbf{w}_g(s), y) : (x, y) \in \mathfrak{D}, s \in S(x)\}.$$

Последующая классификация производится голосованием:

$$\hat{y} = \arg \max_y \sum_{s \in S(x)} 1[b(\mathbf{w}_g(s)) = y].$$

Алгоритм классификации в пространстве гиперпараметров (распределений параметров аппроксимирующих моделей) был описан в [разд. 5 \(Распределения признаков сегментов\)](#). В эксперименте использовалось семейство нормальных распределений с диагональной ковариационной матрицей.

Задача многоклассовой классификации решалась при помощи подхода One-vs-One бинарными классификаторами SVM с RBF ядром и параметрами $C = 100$, $\gamma = 0.017$.

На графике [6](#) ниже приведены результаты для средней точности решения задачи многоклассовой классификации обоими алгоритмами.

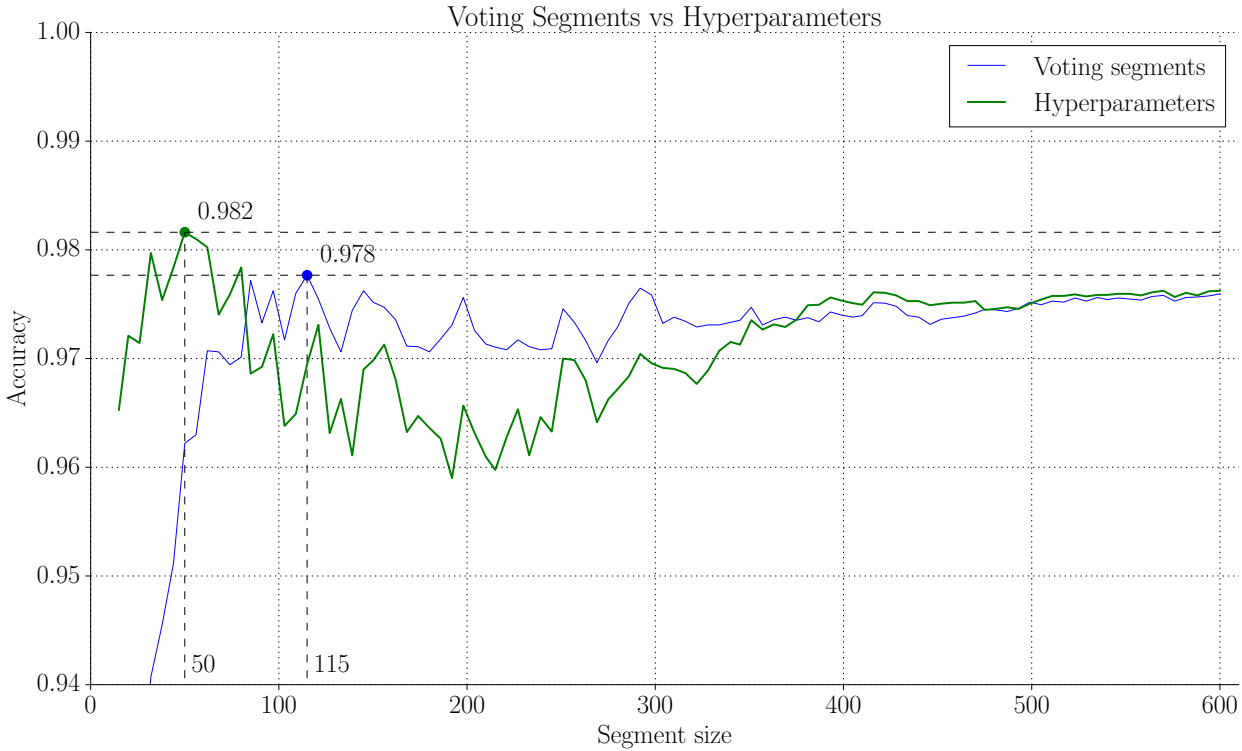


Рис. 6. Зависимость средней точности классификации от длины сегментов. Dataset: USC-HAD, учитываются только первые 10 классов. Под Ассурасу понимается значение функционала (7).

Из графика можно видеть, что оба алгоритма позволяют повысить качество классификации, причем алгоритм классификации в пространстве гиперпараметров при длине сегмента 50 достигает качества 98.2% и показывает результат выше, чем алгоритм голосования.

Объединим результаты из последних двух экспериментов. Будем обучать два классификатора. Первый классификатор a_1 — One-vs-One SVM с RBF ядром и параметрами $C = 10$, $\gamma = 0.13$ — будет разделять классы 11, 12 и первые десять классов для исходных временных рядов. Второй классификатор a_2 — One-vs-One SVM с RBF ядром и параметрами $C = 100$, $\gamma = 0.017$ — классификатор в пространстве гиперпараметров, описанный в предыдущем эксперименте.

Итоговый классификатор выглядит следующим образом:

$$a(x) = \begin{cases} a_1(x), & a_1(x) \in \{11, 12\}, \\ a_2(x), & \text{иначе.} \end{cases} \quad (20)$$

Результаты. На диаграмме 7 ниже приводится качество классификации построенного классификатора 20, усредненное по $r = 500$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

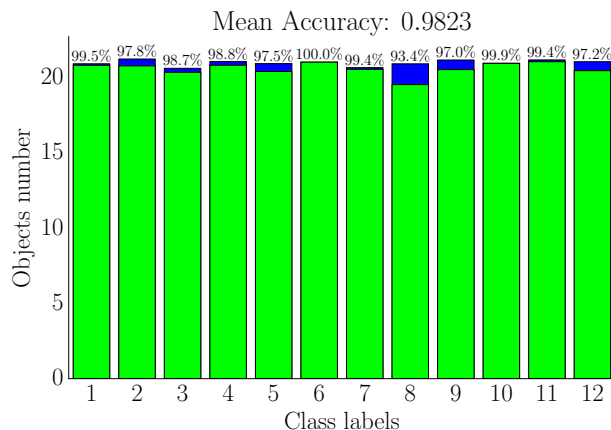


Рис. 7. Dataset: USC-HAD.

		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.01	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.01	0.01	0.00	0.00	0.97	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.06	0.00	0.00	0.00
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.97	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.97

Таблица 7. Mean confusion matrix. Признаки, порожденные моделью авторегрессии. Dataset: USC-HAD.

8 Заключение

В работе было показано, что метод признакового описания временного ряда оптимальными параметрами его аппроксимирующих моделей дает высокое качество решения задачи классификации. Предложенный метод вычислительно эффективен и не требователен к памяти вычислительного устройства. В работе также был предложен алгоритм классификации временных рядов в пространстве распределений параметров порождающих их сегменты моделей, который обобщает предыдущий метод классификации временных рядов и позволяет производить более тонкую настройку алгоритма классификации.

9 Литература

- [1] Tsay, R. S. Analysis of financial time series / R. S. Tsay. — John Wiley & Sons, 2005. — Vol. 543.
- [2] Weiss, G. M. Mining with rarity: a unifying framework / G. M. Weiss // *ACM SIGKDD Explorations Newsletter*. — 2004. — Vol. 6, no. 1. — Pp. 7–19.
- [3] Geurts, P. Segment and combine approach for non-parametric time-series classification / P. Geurts, L. Wehenkel // *Knowledge Discovery in Databases: PKDD 2005*. — Springer, 2005. — Pp. 478–485.
- [4] Liao, T. W. Clustering of time series data—a survey / T. W. Liao // *Pattern Recognition*. — 2005. — Vol. 38, no. 11. — Pp. 1857 – 1874. <http://www.sciencedirect.com/science/article/pii/S0031320305001305>.
- [5] Esling, P. Time-series data mining / P. Esling, C. Agon // *ACM Comput. Surv.* — 2012. — December. — Vol. 45, no. 1. — Pp. 12:1–12:34. <http://doi.acm.org/10.1145/2379776.2379788>.
- [6] Fu, T.-c. A review on time series data mining / T.-c. Fu // *Engineering Applications of Artificial Intelligence*. — 2011. — Vol. 24, no. 1. — Pp. 164–181.

- [7] *Basil, T.* Automatic classification of heartbeats / T. Basil, C. Lakshminarayan // Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European. — 2014. — Sept. — Pp. 1542–1546.
- [8] *Marcel, S.* Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation / S. Marcel, J. d. R. Millán // *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* — 2007. — Vol. 29, no. 4. — Pp. 743–752.
- [9] *Alomari, M. H.* Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning / M. H. Alomari, A. Samaha, K. AlKamha // *International Journal of Advanced Computer Science and Applications.* — 2013. — Vol. 4, no. 6. — Pp. 207–212.
- [10] Human activity recognition using smart phone embedded sensors: A linear dynamical systems method / W. Wang, H. Liu, L. Yu, F. Sun // Neural Networks (IJCNN), 2014 International Joint Conference on. — 2014. — July. — Pp. 1185–1190.
- [11] *Kwapisz, J. R.* Activity recognition using cell phone accelerometers / J. R. Kwapisz, G. M. Weiss, S. A. Moore // *SIGKDD Explor. Newsl.* — 2011. — March. — Vol. 12, no. 2. — Pp. 74–82. <http://doi.acm.org/10.1145/1964897.1964918>.
- [12] *Gruber, C.* Signature verification with dynamic rbf networks and time series motifs / C. Gruber, M. Coduro, B. Sick // Tenth International Workshop on Frontiers in Handwriting Recognition / Suvisoft. — 2006.
- [13] Querying and mining of time series data: Experimental comparison of representations and distance measures / H. Ding, G. Trajcevski, P. Scheuermann et al. // *Proc. VLDB Endow.* — 2008. — August. — Vol. 1, no. 2. — Pp. 1542–1552. <http://dx.doi.org/10.14778/1454159.1454226>.
- [14] *Jeong, Y.-S.* Weighted dynamic time warping for time series classification / Y.-S. Jeong, M. K. Jeong, O. A. Omitaomu // *Pattern Recognition.* — 2011. — Vol. 44, no. 9. — Pp. 2231 – 2240. — Computer Analysis of Images and Patterns. <http://www.sciencedirect.com/science/article/pii/S003132031000484X>.
- [15] *Nanopoulos, A.* Feature-based classification of time-series data / A. Nanopoulos, R. Alcock, Y. Manolopoulos // *International Journal of Computer Research.* — 2001. — Vol. 10. — Pp. 49–61.
- [16] *Wiens, J.* Patient risk stratification for hospital-associated c. diff as a time-series classification task / J. Wiens, E. Horvitz, J. V. Guttag // Advances in Neural Information Processing Systems. — 2012. — Pp. 467–475.
- [17] *Mörchen, F.* Time series feature extraction for data mining using dwt and dft. — 2003.
- [18] *Zhang, H.* A non-parametric wavelet feature extractor for time series classification / H. Zhang, T. B. Ho, M. S. Lin // Advances in Knowledge Discovery and Data Mining. — Springer, 2004. — Pp. 595–603.
- [19] *Corduas, M.* Time series clustering and classification by the autoregressive metric / M. Corduas, D. Piccolo // *Computational Statistics & Data Analysis.* — 2008. — Vol. 52, no. 4. — Pp. 1860 – 1872. <http://www.sciencedirect.com/science/article/pii/S0167947307002368>.
- [20] *Kini, B. V.* Large margin mixture of ar models for time series classification / B. V. Kini, C. C. Sekhar // *Applied Soft Computing.* — 2013. — Vol. 13, no. 1. — Pp. 361–371.
- [21] *Kuznetsov, M. P.* Time series classification algorithm using combined feature description / M. P. Kuznetsov, N. P. Ivkin // *Journal of Machine Learning and Data Analysis.* — 2014. — Vol. 1, no. 11. — Pp. 1471–1483.
- [22] *Motrenko, A. P.* Extracting fundamental periods to segment human motion time series / A. P. Motrenko, V. V. Strijov // *Journal of Biomedical and Health Informatics.* — 2015. — Vol. ? — P. ? http://strijov.com/papers/Motrenko2014TSS_eng.pdf.

- [23] *Vasko, K.* Estimating the number of segments in time series data using permutation tests / K. Vasko, H. Toivonen // *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on.* — 2002. — Pp. 466–473.
- [24] *Ignatov, A.* Human activity recognition using quasiperiodic time series collected from a single triaxial accelerometer / A. Ignatov, V. Strijov // *Multimedia Tools and Applications.* — 2015. — Vol. ? — P. ? <http://strijov.com/papers/Stenina2014RailRoadsMatching.pdf/Ignatov2015HumanActivity.pdf>.
- [25] *Allwein, E. L.* Reducing multiclass to binary: A unifying approach for margin classifiers / E. L. Allwein, R. E. Schapire, Y. Singer // *Journal of Machine Learning Research.* — 2000. — Vol. 1. — Pp. 113–141.
- [26] *Zhang, M.* Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors / M. Zhang, A. A. Sawchuk // *ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware).* — Pittsburgh, Pennsylvania, USA: 2012. — September.