

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, дискретное вейвлет-преобразование.*

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, 2], обнаружение аномалий [3], сегментация [4], кластеризация [5, 6] и классификация [7, 8, 4]. Обзор по задачам и методам анализа временных рядов дается в [9, 10]. Последние годы связаны с ростом интереса к данной области, проявляющимся в непрекращающемся предложении новых методов анализа временных рядов — метрик [11, 12, 13], алгоритмов сегментации [14, 9, 10], кластеризации [15, 16, 9, 10] и других.

В данной работе рассматривается задача классификации временных рядов, возникающая во многих приложениях (медицинская диагностика по ЭКГ [17] и ЭЭГ [18, 19],

классификация типов физической активности по данным с акселерометра [20, 21], верификация динамических подписей [22] и т. д.).

Формально задача классификации в общем виде может быть поставлена 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 алгоритм $a : X \rightarrow Y$ — 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 — функции расстояния [11, 12, 13] на парах временных рядов позволяет применять метрические методы классификации. При удачном выборе метрики дальнейшая классификация может происходить при помощи простейших метрических алгоритмов классификации, например, методом ближайшего соседа [23]. 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, а также позволяет использовать graph-based методы частичного обучения [24, 25].

Другой 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 [26] 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 [27]), а сам выбор признаков становится важной технической задачей. Второй 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 [28] в качестве признаков предлагается использовать коэффициенты дискретного преобразования Фурье (DFT), в [28, 29] — дискретного вейвлет-преобразования (DWT), а в [16, 30, 31] модели авторегрессии. В [32] исследуются свойства смеси моделей авторегрессии. Таким образом, при данном методе построения признаковых описаний возникает задача выбора аппроксимирующей модели временного ряда. Об исчерпывающих исследованиях этой задачи авторам неизвестно.

В работе исследуются методы классификации временных рядов, использующие в качестве их признаковых описаний параметры аппроксимирующих моделей. Приводится сравнение моделей аппроксимации. Как из всякой последовательности, из временного ряда могут извлекаться его подпоследовательности, для которых могут строиться признаковые описания так же, как и для исходных временных рядов. Использование подпоследовательностей (фрагментов) позволяет обобщить алгоритмы классификации. Так в работе [4] 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 распределений признаков их сегментов, который сравнивается с родственным ему алгоритмом голосования сегментов [4]. 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.

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

$$a = b \circ \mathbf{f} \circ S, \quad (1)$$

в которых

— S — процедура сегментации:

$$S : x \mapsto 2^{\mathbf{S}(x)}, \quad (2)$$

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

- \mathbf{f} — процедура построения признаковового описания набора сегментов,
- b — алгоритм многоклассовой классификации.

Задана функция потерь

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

и функционал качества

$$Q(a, \mathfrak{D}) = \frac{1}{|\mathfrak{D}|} \sum_{(x,y) \in \mathfrak{D}} \mathcal{L}(x, a(x), y). \quad (3)$$

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

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

где b^* — минимизатор эмпирического риска:

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

Оптимальный метод обучения определяется по скользящему контролю:

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

3 Сегментация временных рядов

Определение 1. Фрагментом временного ряда $x = [x^{(1)}, \dots, x^{(t)}]$ будем называть любую его подпоследовательность $s = [x^{(t_1)}, \dots, x^{(t_k)}]$, где $1 \leq t_1 < \dots < t_k \leq t$.

Определение 2. Сегментом временного ряда x будем называть его непрерывный фрагмент $s = [x^{(i)}]_{i=t_0}^{t_1}$, $1 \leq t_0 \leq t_1 \leq t$.

Определение 3. Под сегментацией будем понимать отображение временных рядов во множество их сегментов (*выр. 2*).

Примеры.

— Тривиальная сегментация

$$S : x \mapsto \{x\}. \quad (4)$$

— Случайное выделение сегментов некоторой длины ℓ [4].

— Важным является случай квазипериодичности временного ряда, когда сам ряд состоит из похожих в определенном смысле сегментов, называемых периодами:

$$x = \left[\underbrace{x^{(1)}, \dots, x^{(t_1)}}_{s^{(1)}}, \underbrace{x^{(t_1+1)}, \dots, x^{(t_2)}}_{s^{(2)}}, \dots, \underbrace{x^{(t_{p-1}+1)}, \dots, x^{(t)}}_{s^{(p)}} \right]. \quad (5)$$

Тогда в качестве процедуры сегментации можно взять разбиение на периоды:

$$S : x \mapsto \{s^{(1)}, \dots, s^{(p)}\}. \quad (6)$$

Для выделения периодов могут быть использованы, например, алгоритмы [33, 14, 34].

4 Аппроксимирующая модель сегмента временного ряда

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

Определение 4. Параметрической аппроксимирующей моделью временного ряда x будем называть отображение

$$g : \mathbb{R}^n \times X \rightarrow X. \quad (7)$$

В слово «аппроксимирующая» вкладывается тот смысл, что модель должна приближать временной ряд:

$$\rho(g(\mathbf{w}, x), x) < \varepsilon \text{ для некоторого } \mathbf{w} \in \mathbb{R}^n.$$

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

Определение 5. Признаковым описанием объекта x , порожденным параметрической моделью $g(\mathbf{w}, x)$ назовем вектор оптимальных параметров этой модели

$$\mathbf{f}_g(x) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \rho(g(\mathbf{w}, x), x). \quad (8)$$

Приведем несколько примеров.

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

$$x = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)}], \text{ где } \mathbf{x}^{(k)} = [x_0^{(k)}, \dots, x_r^{(k)}]^\top, \quad k = 1, \dots, t.$$

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

$$g(\mathbf{w}, x) = [\hat{\mathbf{x}}^{(1)}, \dots, \hat{\mathbf{x}}^{(t)}], \text{ где } \hat{\mathbf{x}}^{(k)} = [\hat{x}_0^{(k)}, x_1^{(k)}, \dots, x_r^{(k)}]^\top, \quad k = 1, \dots, t,$$

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} \hat{x}_0^{(1)} \\ \vdots \\ \hat{x}_0^{(t)} \end{bmatrix} = \underbrace{\begin{bmatrix} x_1^{(1)} & \dots & x_r^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(t)} & \dots & x_r^{(t)} \end{bmatrix}}_X \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_r \end{bmatrix}}_{\mathbf{w}}.$$

Тогда, выбрав в качестве ρ евклидово расстояние, по определению 5 получим признаковое описание объекта x :

$$\mathbf{f}_g(x) = (X^\top X)^{-1} X^\top \hat{\mathbf{x}}_0.$$

- **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 :

$$g(\mathbf{w}, x) = [\hat{x}^{(1)}, \dots, \hat{x}^{(t)}], \text{ где } \hat{x}^{(k)} = w_0 + \sum_{i=1}^p w_i x^{(k-i)}, \quad k = 1, \dots, t. \quad (9)$$

Далее признаковое описание определяется аналогично случаю линейной регрессии.

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

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

Взяв в качестве аппроксимирующей модели обратное преобразование Фурье,

$$g(\mathbf{w}, x) = [\hat{x}^{(0)}, \dots, \hat{x}^{(t-1)}], \text{ где } \hat{x}^{(k)} = \frac{1}{t} \sum_{j=0}^{t-1} w_j e^{\frac{2\pi i}{t} k j}, \quad k = 0, \dots, t-1,$$

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

$$\mathbf{f}_g(x) = [\operatorname{Re} w_0, \operatorname{Im} w_0, \dots, \operatorname{Re} w_{t-1}, \operatorname{Im} w_{t-1}], \text{ где } w_k = \sum_{j=0}^{t-1} x^{(j)} e^{-\frac{2\pi i}{t} k j}, \quad k = 0, \dots, t-1. \quad (10)$$

Приведенные примеры демонстрируют большую общность построения пространства признаков при помощи моделей типа 7 и решения оптимизационной задачи 8. Вообще говоря, легко видеть, что при $|X| \geq 2$ любая процедура построения признаков описаний $\mathbf{f} : X \rightarrow \mathbb{R}^n$ задается решением оптимизационной задачи 8 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 [7](#) получим для каждого сегмента $s^{(k)} \in S(x) = \{s^{(1)}, \dots, s^{(p)}\}$ временного ряда x его признаковое описание $\mathbf{f}^{(k)} := \mathbf{f}_g(s^{(k)})$, решив оптимизационную задачу [8](#). Тогда всему набору сегментов $S(x)$ будет соответствовать выборка

$$\left(\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}\right). \quad (11)$$

Будем рассматривать вероятностную модель.

Гипотеза 1. Выборка $\left(\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}\right)$ — простая, то есть случайная, независимая и однородная, где $\mathbf{f}^{(k)} \sim P_0$.

Далее будем рассматривать вопрос оценки распределения P_0 .

Пусть имеется параметрическое семейство распределений $\{P_\theta\}_{\theta \in \Theta}$. Тогда, получив оценку максимального правдоподобия

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}),$$

будем считать её признаковым описанием исходного временного ряда. Таким образом, задача классификации временных рядов свелась к задаче классификации параметров распределений из семейства $\{P_\theta\}_{\theta \in \Theta}$. При этом, имея априорное распределение параметра $F(\theta)$, целесообразно использовать вместо оценки максимального правдоподобия апостериорное математическое ожидание параметра

$$\hat{\theta} = E[\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}] = \int_{\Theta} \theta dF(\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}),$$

минимизирующее средний квадрат отклонения от истинного значения параметра, где

$$dF(\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}) = \frac{\mathcal{L}(\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}) dF(\theta)}{\int_{\Theta} \mathcal{L}(\theta | \mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)}) dF(\theta)}.$$

Заметим, что в частном случае тривиальной сегментации [4](#) и семейства вырожденных распределений оценка $\hat{\theta}$ является исходным признаковым описанием. Таким образом, предложенный подход к построению признакового описания временного ряда

$$\mathbf{f} : x \mapsto \hat{\theta} \quad (12)$$

является достаточно общим и при этом хорошо интерпретируется.

6 Алгоритм классификации

Для завершения построения алгоритма 1 осталось рассмотреть алгоритмы многоклассовой классификации b . К этому моменту мы уже имеем признаковое описание $\mathbf{f}(x)$ для каждого временного ряда x обучающей выборки $\mathfrak{D} \subset X \times Y$.

Решать задачу многоклассовой классификации будем сведением её к задачам бинарной классификации. Наиболее общей стратегией к сведению является Error-Correcting Output Codes [35], 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 [21] содержит показания акселерометра для 6 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. Каждой компоненте временного ряда сопоставлялись 40 чисел — её среднее, стандартное отклонение, средний модуль отклонения от среднего, гистограмма 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 гауссовским ядром и параметрами $C = 8.5$, $\gamma = 0.12$.

Результаты. Для оценки качества решения использовалась процедура скользящего контроля. Исходная обучающая выборка \mathfrak{D} случайно разбивается m раз на обучающую и контрольную ($\mathfrak{D} = \mathfrak{L}_i \sqcup \mathfrak{T}_i$). В качестве внешнего критерия качества метода обучения μ бралось

$$\frac{1}{m} \sum_{i=1}^m Q(\mu(\mathfrak{L}_i), \mathfrak{T}_i),$$

где для средней точности (ассигасы) классификации объектов класса $c \in Y$

$$Q_c(a, \mathfrak{T}) = \frac{\sum_{\substack{(x,y) \in \mathfrak{T} \\ y=c}} \mathbf{1}\{a(x) = y\}}{\sum_{\substack{(x,y) \in \mathfrak{T} \\ y=c}} 1}, \quad (13)$$

а для среднего качества решения задачи многоклассовой классификации

$$Q(a, \mathfrak{T}) = \frac{1}{|\mathfrak{T}|} \sum_{(x,y) \in \mathfrak{T}} \mathbf{1}\{a(x) = y\}. \quad (14)$$

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

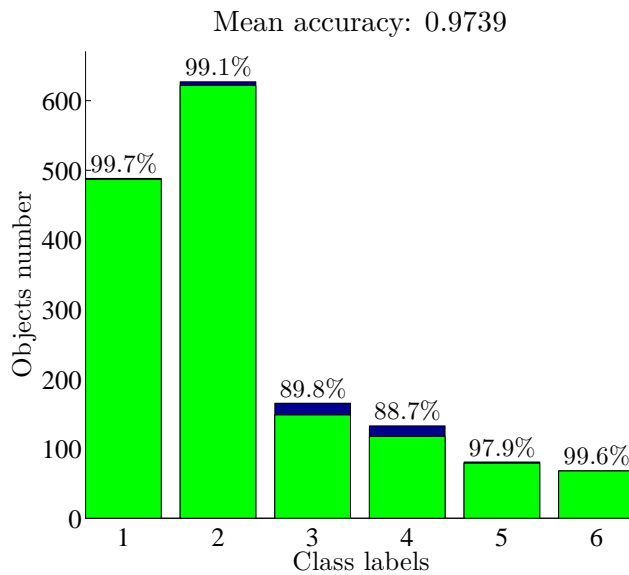


Рис. 1. Точность классификации при ручном выделении признаков. Dataset: WISDM. Под Mean ассурасу понимается значение функционала 14. Над столбцами приведены средние точности классификации для каждого класса по формуле 13.

Как видно из таблицы 2, классы 2, 3 и 4 не достаточно 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	0.90	0.04	0.00	0.00
	4	0.01	0.05	0.05	0.89	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 9

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. 9). Таким образом, каждый временной ряд описывался 31 числами. Так же, проводилась предварительная нормализация признаков.

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

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

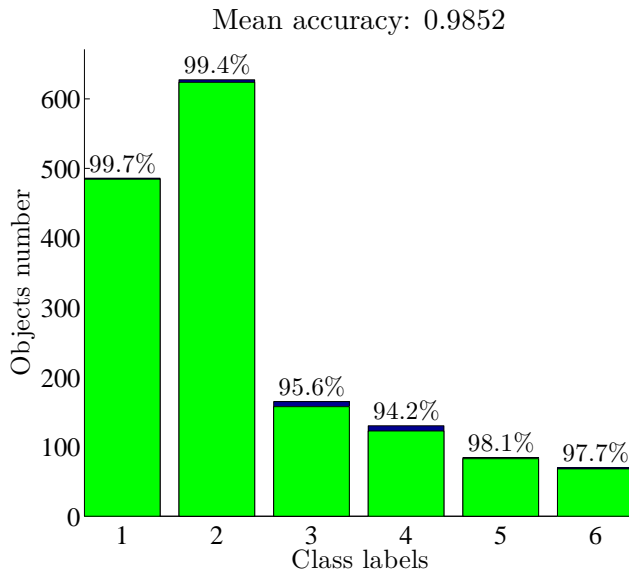


Рис. 2. Точность классификации для параметров модели авторегрессии в качестве признаков описаний. Dataset: WISDM. Под Mean accuracy понимается значение функционала 14. Над столбцами приведены средние точности классификации для каждого класса по формуле 13.

Таким образом, несмотря на неравномерное распределение объектов по классам, использование признаков описания, порожденного моделью авторегрессии, позволяет значительно повысить качество классификации. Точность построенного классификатора минимальна для 4 класса «Downstairs» и составляет 94.2%. Алгоритм классификации можно улучшать добавлением дополнительных признаков, например, параметров линейной регрессии, однако, добиваться максимально возможной точности не входит в цели нашего эксперимента.

7.2 Датасет USC-HAD

Датасет USC-HAD [36] 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 рядов для каждого класса, а средняя длина временного ряда порядка 3300. Частота записи измерений сенсора 100 Hz.

		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	0.01	0.96	0.02	0.00	0.00
	4	0.00	0.02	0.04	0.94	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.98	0.01
	6	0.01	0.00	0.00	0.00	0.01	0.98

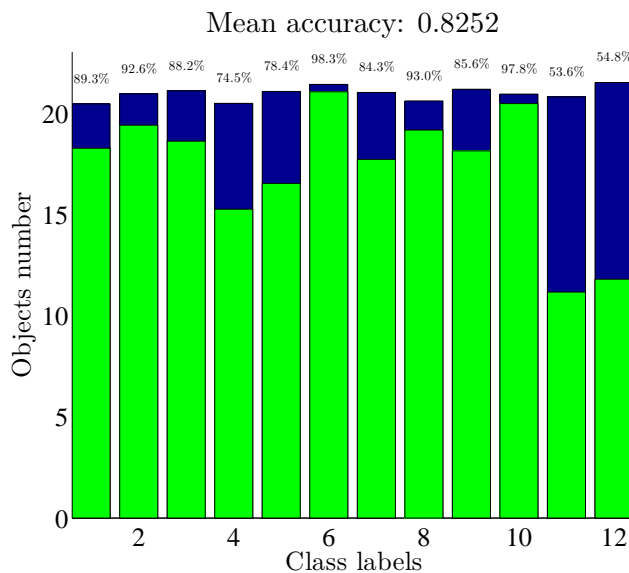
Таблица 3. Mean confusion matrix. Признаки, порожденные моделью авторегрессии. Dataset: WISDM.

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 гауссовским 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	0.89	0.00	0.00	0.04	0.03	0.01	0.03	0.00	0.00	0.00	0.00	0.00
	2	0.00	0.93	0.01	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.01	0.88	0.07	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.01	0.10	0.74	0.08	0.01	0.03	0.00	0.00	0.00	0.00	0.00
	5	0.05	0.06	0.00	0.04	0.78	0.01	0.06	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00
	7	0.01	0.00	0.00	0.01	0.05	0.09	0.84	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	0.03
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.86	0.00	0.07	0.05
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.98	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	0.54	0.43
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.40	0.55

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

Рис. 3. Точность классификации при ручном выделении признаков. Dataset: USC-HAD. Под Mean ассигасу понимается значение функционала [14](#). Над столбцами приведены средние точности классификации для каждого класса по формуле [13](#).

Из таблицы [4](#) видно, что классы 11 и 12 (elevator up и elevator down) очень 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%.

7.2.2 Модель авторегрессии [9](#)

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 возьмем 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 добавим по 11 параметров авторегрессионной модели AR(10) (см. [выр. 9](#)). Затем проведем нормализация признаков.

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

Результаты. На диаграмме ниже приведено качество классификации, усредненное по $m = 200$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

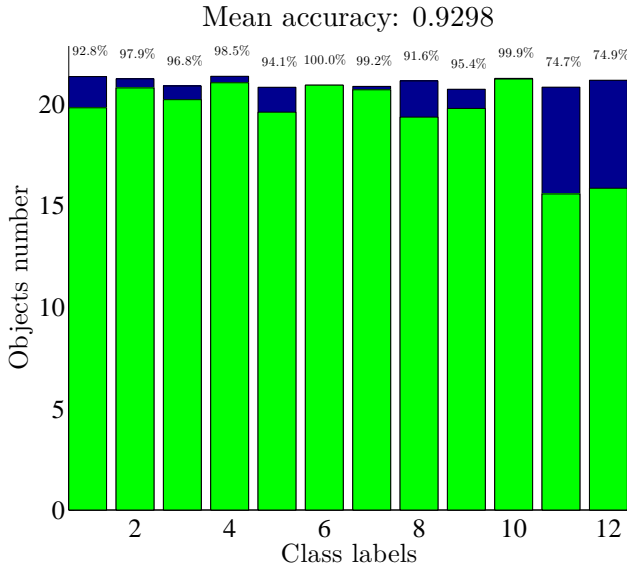


Рис. 4. Точность классификации для параметров модели авторегрессии в качестве признаков описаний. Dataset: USC-HAD. Под Mean accuracy понимается значение функционала [14](#). Над столбцами приведены средние точности классификации для каждого класса по формуле [13](#).

Из [таблицы 5](#) видно, что использование признаков описания, порожденного моделью авторегрессии, значительно повысило качество классификации для всех классов. Недостаточно точно отделяются только классы 11 и 12 (elevator up и elevator down), где ошибка составляет 25%. Однако, они превосходно отделяются от всех остальных классов. Поэтому, для достижения приемлемого качества классификации имеет смысл рассматривать эти классы отдельно и выбирать бинарный классификатор независимо от остальных классов.

		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	0.93	0.01	0.03	0.02	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	0.02	0.02	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	0.01	0.94	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	0.00	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	0.06	0.00	0.02	0.00
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.95	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	0.75	0.25	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.22	0.75

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

7.2.3 Модель авторегрессии [9](#) и Фурье [10](#)

Признаковое описание. Возьмем признаковое описание временных рядов из предыдущего эксперимента, [разд. 7.2.2 \(Признаковое описание\)](#), и добавим к нему первые 5 коэффициентов Фурье [10](#). Таким образом, каждый 6-компонентный временной ряд будет описываться 128 признаками.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One и квадратичной функции потерь. В качестве бинарного классификатора использовался SVM с гауссовским ядром и параметрами $C = 16$, $\gamma = 0.04$.

Результаты. На диаграмме ниже приведено качество классификации, усредненное по $m = 100$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

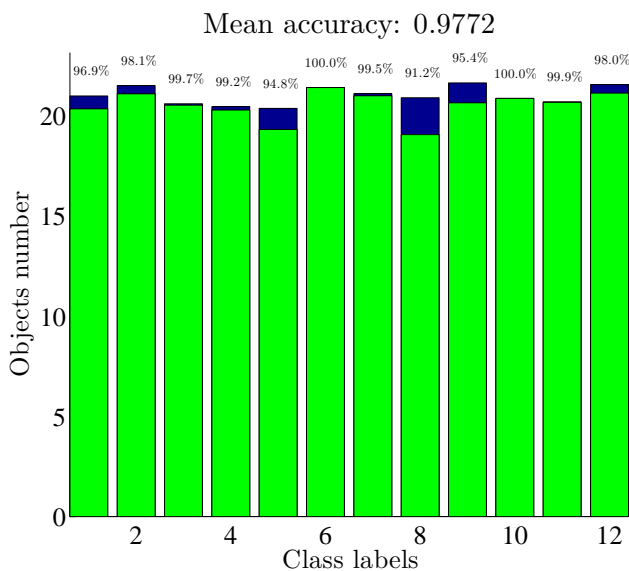


Рис. 5. Точность классификации для параметров модели авторегрессии в качестве признаковых описаний. Dataset: USC-HAD. Под Mean accuracy понимается значение функционала [14](#). Над столбцами приведены средние точности классификации для каждого класса по формуле [13](#).

Из [таблицы 6](#) видно, что использование коэффициентов Фурье значительно повысило качество классификации. Классы 8 и 9 (sit and fidget и stand) достоверно отделяются от всех остальных десяти классов. Поэтому повысить точность для них можно, настраивая отдельно на этих двух классах бинарный классификатор, их разделяющий в подходе One-vs-One.

		Predicted class											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual class	1	0.97	0.01	0.01	0.01	0.01	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	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	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.00	0.00	0.01	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	0.00	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	0.91	0.08	0.00	0.00	0.01
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.95	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	0.01	0.00	0.00	0.01	0.98

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

7.2.4 Классификация голосованием и классификация в пространстве распределений параметров

Рассмотрим, наконец, алгоритм классификации в сочетании с процедурой сегментации временных рядов. В качестве процедуры сегментации $S(x)$ (см. [2](#)) будем использовать

случайное выделение сегментов переменной длины. Для каждого временного ряда получаем 20 сегментов. Будем решать задачу классификации только 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 $a(x)$ обучается на новой обучающей выборке для сегментов исходных временных рядов

$$\hat{\mathcal{D}} = \{(\mathbf{f}_g(s), y) : (x, y) \in \mathcal{D}, s \in S(x)\}.$$

А последующая классификация производится голосованием:

$$\text{mode} \{a(\mathbf{f}_g(s)) : s \in S(x)\}.$$

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 линейным ядром и параметром $C = 0.25$.

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

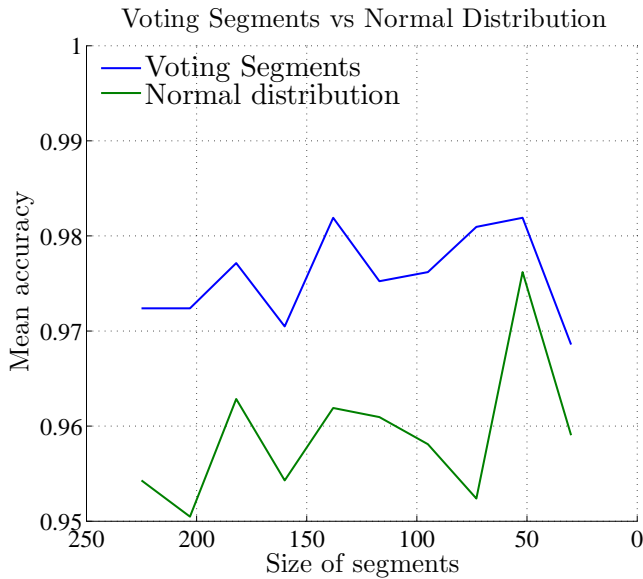


Рис. 6. Зависимость средней точности классификации от длины случайных сегментов. Dataset: USC-HAD, учитываются только первые 10 классов. Под Mean ассигасу понимается значение функционала [14](#).

Из графика можно видеть, что оба алгоритма позволяют повысить точность классификации на 1% при длине сегмента 50. Однако, при любой длине сегмента алгоритм голосования показывает лучшие результаты.

8 Обсуждение

Сравнение результатов

9 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. Однако, вычислительный эксперимент на данных задачи классификации типов физической активности показал, что в сочетании с процедурой случайной сегментации алгоритм голосования сегментов позволяет добиться лучших результатов.

10 Литература

- [1] *Weigend, A. S.* Time series prediction: forecasting the future and understanding the past / A. S. Weigend // *Santa Fe Institute Studies in the Sciences of Complexity*. — 1994.
- [2] *Tsay, R. S.* Analysis of financial time series / R. S. Tsay. — John Wiley & Sons, 2005. — Vol. 543.
- [3] *Weiss, G. M.* Mining with rarity: a unifying framework / G. M. Weiss // *ACM SIGKDD Explorations Newsletter*. — 2004. — Vol. 6, no. 1. — Pp. 7–19.
- [4] *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.
- [5] *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>.
- [6] *Zolhavarieh, S.* A review of subsequence time series clustering / S. Zolhavarieh, S. Aghabozorgi, Y. W. Teh // *The Scientific World Journal*. — 2014. — Vol. 2014.
- [7] 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 / IEEE. — 2014. — Pp. 1185–1190.
- [8] *Wei, L.* Semi-supervised time series classification / L. Wei, E. Keogh // *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. — KDD '06. — New York, NY, USA: ACM, 2006. — Pp. 748–753. <http://doi.acm.org/10.1145/1150402.1150498>.
- [9] *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>.
- [10] *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.

- [11] 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>.
- [12] Salvador, S. Toward accurate dynamic time warping in linear time and space / S. Salvador, P. Chan // *Intelligent Data Analysis.* — 2007. — Vol. 11, no. 5. — Pp. 561–580.
- [13] Marteau, P.-F. Time warp edit distance with stiffness adjustment for time series matching / P.-F. Marteau // *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* — 2009. — Vol. 31, no. 2. — Pp. 306–318.
- [14] 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.
- [15] Fröhwrth-Schnatter, S. Model-based clustering of multiple time series / S. Fröhwrth-Schnatter, S. Kaufmann // *Journal of Business & Economic Statistics.* — 2008. — Vol. 26, no. 1. — Pp. 78–89.
- [16] 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>.
- [17] 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.
- [18] 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.
- [19] 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 // *arXiv preprint arXiv:1312.2877.* — 2013.
- [20] 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.
- [21] 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>.
- [22] 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.
- [23] 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>.
- [24] Nguyen, M. N. Positive unlabeled learning for time series classification. / M. N. Nguyen, X.-L. Li, S.-K. Ng // *IJCAI / Citeseer.* — Vol. 11. — 2011. — Pp. 1421–1426.
- [25] Marussy, K. Success: a new approach for semi-supervised classification of time-series / K. Marussy, K. Buza // *Artificial Intelligence and Soft Computing / Springer.* — 2013. — Pp. 437–447.
- [26] 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.

- [27] *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.
- [28] *Mörchen, F.* Time series feature extraction for data mining using dwf and dft. — 2003.
- [29] *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.
- [30] *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.
- [31] *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.
- [32] *Kalliovirta, L.* A gaussian mixture autoregressive model for univariate time series / L. Kalliovirta, M. Meitz, P. Saikkonen // *Journal of Time Series Analysis*. — 2015. — Vol. 36, no. 2. — Pp. 247–266.
- [33] *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.
- [34] *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](http://strijov.com/papers/Ignatov2015HumanActivity.pdf).
- [35] *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.
- [36] *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.