

Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию*

М. П. Кузнецов, Н. П. Ивкин

mikhail.kuznecov@phystech.edu, ivkinnikita@gmail.com

Московский физико-технический институт, Москва

Рассматривается задача многоклассовой классификации временных рядов. Временные ряды являются объектами сложной структуры, для которых не задано исходное признаковое описание. Исследуются различные методы построения признакового пространства для временного ряда: метод экспертного задания порождающих функций и метод построения признаков на основе гипотезы порождения данных. Рассматривается комбинированное признаковое описание временного ряда. В качестве прикладной задачи рассматривается задача классификации данных акселерометра. Показано, что использование расширенного множества признаков приводит к значительному улучшению качества классификации.

Ключевые слова: *временные ряды; авторегрессионная модель; сингулярный спектр; метод опорных векторов*

Time series classification algorithm using combined feature description*

M. P. Kuznetsov and N. P. Ivkin

Moscow Institute of Physics and Technology, Moscow, Russia

A problem of time series multiclass classification is considered. Time series are regarded as complex-structured objects having no explicit feature description. In general, complex objects classification problem can be divided into two stages: to form feature space and to construct a decision rule on this space. The focus of the paper is on the first stage, namely, to construct the feature space where the points of different classes are separable or close to separable. Having constructed such space, a simple linear or polynomial decision rule is used to discriminate the classes.

Various methods of time series feature space construction are investigated. The first method is the expert definition of basic functions. Namely, for the time series data, mean value, deviation, absolute deviation, and empirical distribution of values are considered. The second method involves data generation hypothesis and uses optimal estimations of generation parameters as the considered features. Furthermore, a combined feature description of a time series is considered. The computations show that using the extended feature space allows to significantly improve the classification quality.

The proposed approach is used for the accelerometer time series classification. The problem is to classify each time series segment to one of six classes-actions: Jogging, Walking, Upstairs, Downstairs, Sitting, and Standing. It is shown that the combined approach achieves very good accuracy comparing with the separate feature construction methods.

Keywords: *time series; autoregressive model; singular spectrum; support vector machine*

*Работа выполнена при финансовой поддержке РФФИ, проект 15-37-50324 мол_нр.

1 Введение

Решается задача классификации объектов сложной структуры, т.е. таких объектов, для которых не сформировано исходное признаковое описание и матрица плана. Объекты подобного типа возникают во многих задачах анализа данных: распознавание объектов на изображении, классификация звуковых сигналов, тематическое моделирование. В данной работе рассматривается задача классификация одномерных и многомерных временных рядов произвольной, возможно, различной длины [1, 2]. В этих терминах временной ряд является объектом сложной структуры: значения временного ряда, соответствующие различным временным отсчетам, не могут рассматриваться в качестве признакового описания объекта.

В качестве прикладной задачи классификации временных рядов рассматривается задача классификация данных акселерометра [3]. Данные представляют собой измерения акселерометра некоторого устройства, например мобильного телефона в кармане человека, и могут использоваться для идентификации действия человека в каждый момент времени. Примером действия служат ходьба, бег, подъем по лестнице и др. Ранее задача распознавания данных акселерометра для определения типа действия человека также была поставлена в форме задачи онлайн сегментации временного ряда [4]. В отличие от задачи сегментации в данной работе исследуются предобработанные данные: размеченные временные ряды, соответствующие различным действиям. Решение задачи классификации в такой постановке позволяет выделить некоторые характерные признаки, которые могут быть использованы в дальнейших работах.

В общем случае задача классификации объектов сложной структуры разделяется на два этапа. На первом этапе формируется признаковое описание объектов, на втором этапе строится решающее правило классификации. Отметим, что эти два этапа, вообще говоря, зависимы друг от друга, однако в простейшем случае они могут рассматриваться отдельно. В данной работе основной упор сделан на построение признакового пространства, наиболее полным образом описывающего выделяемые классы действия; классификаторы на втором этапе ищутся в классе линейных или низкой степени полиномиальных решающих правил. Такая мотивация следует из отделяемости этапа построения признакового пространства от задачи классификации. Построенные признаки, если они описывают целевую переменную достаточно адекватно, предлагается использовать в последующих задачах сегментации и онлайн обучения.

В работе рассматриваются два основных метода построения признакового пространства. Первый метод заключается в экспертном назначении базовых функций. В [5] выделяют такие базовые функции, как среднее временного ряда, стандартное отклонение, среднее расстояние между пиками, распределение значений временного ряда. Воспользуемся этим методом в качестве получения базового признакового пространства и результатов классификации.

Второй метод построения признакового пространства заключается в том, что назначается параметрическая гипотеза порождения объекта сложной структуры. В частности, для временного ряда в качестве такой гипотезы рассматривается модель авторегрессии, дополняемая анализом условной гетероскедастичности [6]. Для каждого объекта — временного ряда — вычисляются оптимальные параметры порождения, в частном случае коэффициенты авторегрессии; эти вычисленные параметры составляют новое признаковое пространство [7, 8]. Таким образом, процедура классификации выполняет разбиение пространства параметров модели на области, принадлежащие различным классам. Помимо авторегрес-

сионной модели исследуется также модель сингулярного спектра временного ряда, где выделяемыми признаками являются собственные числа траекторной матрицы [9].

Структура работы организована следующим образом. В первом разделе поставлена задача классификации объектов сложной структуры и выделения признаков в общем смысле. Во втором разделе рассмотрена задача классификации временных рядов, приводится конкретный вид гипотез порождения данных для выделения признаков на основе оценки оптимальных параметров порождения. В третьем разделе рассматривается задача классификации данных акселерометра, описывается процедура ручной генерации признаков. Раздел вычислительного эксперимента выполняет сравнение предлагаемых методов выделения признаков. Оказывается, что различные методы помогают лучше выделять отдельные классы. Окончательный эксперимент использует полное признаковое описание, заключающее в себе все предлагаемые методы выделения признаков.

2 Классификация объектов сложной структуры

2.1 Постановка задачи классификации

Пусть $s \in \mathcal{S}$ — объект сложной структуры. Рассматривается задача восстановления зависимости

$$y = f(s),$$

где функция f отображает пространство объектов сложной структуры \mathcal{S} в пространство ответов Y , $y \in Y$.

Задана выборка \mathcal{D} объектов сложной структуры и ответов:

$$\mathcal{D} = \{(s_i, y_i)\}_{i=1}^m.$$

Задана функция потерь $l(f(s_i), y_i)$, выражающая величину ошибки классификации функции f на объекте s_i выборки \mathcal{D} . Требуется найти функцию f , минимизирующую суммарные потери на выборке \mathcal{D} .

2.2 Пространство признаков объекта сложной структуры

Отображение f будем рассматривать в классе суперпозиций $f = g \circ h$ таких, что

$$f(s) = g(h(s), \mathbf{b}),$$

где

$$h(s) : \mathcal{S} \rightarrow \Theta \subset \mathbb{R}^n$$

является отображением пространства \mathcal{S} в признаковое пространство $\Theta \subset \mathbb{R}^n$, а $g(\mathbf{h}, \mathbf{b})$ является параметрическим отображением пространства Θ в пространство ответов Y .

Через \hat{y}_i обозначим значение функции f на объекте s_i :

$$\hat{y}_i = f(s_i) = g(\mathbf{h}_i, \mathbf{b}),$$

где $h_{ij} \equiv h_j(s_i)$ является j -й компонентой значения вектор-функции h на объекте s_i . Значение h_{ij} будем называть j -м признаком объекта s_i , или j -й статистикой.

В данной работе рассматриваются следующие способы построения признакового пространства h_{ij} :

- способ ручной генерации признаков. Такой способ использует экспертную информацию о структуре сложного объекта. Например, в рассматриваемой задаче классификации временных рядов признаками могут быть среднее значение временного ряда на отрезке, среднеквадратичное отклонение, максимальное значение и другие статистики;

– способ введения гипотезы порождения объекта сложной структуры. В этом случае статистики h_{ij} являются оценками параметров рассматриваемой гипотезы. Подробнее об этом методе будет рассказано в следующем параграфе.

2.3 Гипотеза порождения объекта сложной структуры

Будем полагать, что объект сложной структуры s_i представляет собой множество реализаций объектов «простой структуры»:

$$s_i = \{x_{i1}, \dots, x_{it(i)}\},$$

где x_{it} является реализацией случайной величины $X_i \sim P_{\theta_i}$, $\theta_i \in \Theta \subset \mathbb{R}^n$.

Будем полагать, что задана функция ошибки $S(s_i, \theta, \lambda)$, имеющая одинаковый вид для всех объектов сложной структуры s_i . Эта функция ошибки может являться, например, функцией правдоподобия данных X_i и служит для определения оптимальных параметров $\hat{\theta}_i$ для объекта s_i :

$$\hat{\theta}_i = \arg \min_{\theta \in \Theta} S(s_i, \theta, \lambda). \quad (1)$$

Вектор λ является вектором внешних для функции ошибки S параметров. Будем называть λ вектором *структурных параметров*.

В качестве нового признакового описания h_{ij} объекта s_i будем рассматривать оптимальные значения вектора параметров $\hat{\theta}_i$:

$$\mathbf{h}_i \equiv \hat{\theta}_i(s_i).$$

Таким образом, для задачи классификации осуществляется разбиение пространства параметров Θ порождения объектов сложной структуры s на множества принадлежности меткам классов Y .

Частный вид гипотез порождения данных, структурных параметров и функции ошибки для задачи классификации временных рядов будет приведен в следующих разделах. Например, гипотезой порождения будет служить авторегрессионная модель порождения, структурным параметром — длина авторегрессионной модели, а функцией ошибки — качество прогнозирования временного ряда.

2.4 Определение оптимальных параметров

Согласно вышеизложенному, задача определения оптимальных параметров проводится в два этапа.

1. Построение признакового пространства h_{ij} объекта сложной структуры s_i . Статистики \mathbf{h}_i могут быть построены вручную экспертным путем, либо путем минимизации функции ошибки (1) для гипотезы порождения объекта s_i :

$$\mathbf{h}_i \equiv \hat{\theta}_i = \arg \min_{\theta \in \Theta} S(s_i, \theta, \lambda).$$

2. Определение оптимальных параметров $\hat{\mathbf{b}}$ в задаче классификации объектов $(\mathbf{h}_i, y_i)_{i=1}^m$ выборки \mathfrak{D} в новом признаковом пространстве,

$$\hat{y}_i = g(\mathbf{h}_i, \hat{\mathbf{b}}),$$

путем минимизации функции ошибки:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} L(g, \mathbf{b}, \lambda, \mathfrak{D}). \quad (2)$$

3 Многоклассовая классификация одномерных временных рядов

В качестве частного случая объектов сложной структуры в данной работе будут рассматриваться одномерные (а далее — и многомерные) временные ряды переменной длины. Объектом s_i в данном случае будет являться последовательность

$$s_i = \{x_1, \dots, x_{T(i)}\},$$

где длина временного ряда $T(i)$ является переменной и зависит от индекса i .

Задана выборка $\mathcal{D} = \{(s_i, y_i)\}_{i=1}^m$, где s_i — временной ряд, y_i — метка класса. Будем строить отображение временных рядов в пространство меток классов Y в виде классификации параметров $\hat{\theta}_i$ модели авторегрессии, определенных для каждого из временных рядов s_i :

$$\hat{y}_i = g(\hat{\theta}_i, \hat{\mathbf{b}}),$$

где оптимальные параметры $\hat{\mathbf{b}}$ минимизируют ошибку классификации (2):

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \sum_{i=1}^m I(g(\hat{\theta}_i, \mathbf{b}) \neq y_i). \quad (3)$$

Здесь вид функции g является одним из стандартных методов классификации, в частности в данной работе рассматриваются метод многоклассовой логистической регрессии и метод опорных векторов (SVM — support vector machine).

Далее рассмотрим различные варианты выделения признаков h_{ij} . Будут рассмотрены способы введения гипотезы порождения временного ряда s_i и оценки параметров порождения, $\mathbf{h}_i \equiv \hat{\theta}_i(s_i)$, а также способ ручной генерации признаков на основе экспертно заданных функций.

3.1 Модель авторегрессии для одномерного временного ряда

В качестве гипотезы порождения временного ряда s рассмотрим авторегрессионную модель порядка n . Здесь n является структурным параметром, элементом (в данном случае, единственным) вектора λ :

$$x_t = \theta_0 + \sum_{j=1}^n \theta_j x_{t-j} + \varepsilon_t.$$

В этом случае, оптимальные параметры $\hat{\theta}_i \equiv \mathbf{h}_i$ для объекта s_i определяются минимизацией среднеквадратичной ошибки прогнозирования

$$\hat{\theta}_i = \arg \min_{\theta \in \Theta} S(s_i, \theta, \lambda) = \arg \min_{\theta \in \Theta} \left(\sum_{t=1}^{T(i)} \|x_t - \hat{x}_t\|^2 \right), \quad (4)$$

где

$$\hat{x}_t = \theta_0 + \sum_{j=1}^n \theta_j x_{t-j}.$$

3.2 Анализ сингулярного спектра

В качестве альтернативной гипотезы порождения данных рассмотрим модель ряда SSA (Singular Spectrum Analysis) [9]. Поставим в соответствие временному ряду $s = \{x_1, \dots, x_T\}$ траекторную матрицу \mathbf{X} , т. е. матрицу следующего вида:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m-n+1} & x_{m-n+2} & \cdots & x_m \end{pmatrix}.$$

Построим сингулярное разложение матрицы $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{H}\mathbf{V}^T, \quad \mathbf{H} = \text{diag}(h_1, \dots, h_n),$$

где h_1, \dots, h_n — собственные числа матрицы $\mathbf{X}^T\mathbf{X}$, отвечающие за величины различных частот спектра временного ряда s . Для временного ряда s_i будем рассматривать вектор сингулярных чисел \mathbf{h}_i в качестве нового признакового описания и строить классификатор \hat{y}_i в виде

$$\hat{y}_i = g(\mathbf{h}_i, \mathbf{b}),$$

где оптимальные параметры $\hat{\mathbf{b}}$ минимизируют ошибку (3).

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 в следующем разделе.

4 Классификация данных акселерометра

В качестве прикладной задачи рассматривается задача классификации активности человека по данным с акселерометра. Данные с акселерометра представляют собой трехмерный временной ряд $\{a_x(t), a_y(t), a_z(t)\}_{t=1}^T$. Каждому временному ряду поставлена в соответствие метка класса y_i , обозначающая один из шести возможных типов действия: Jogging (Бег), Walking (Ходьба), Upstairs (Ходьба вверх), Downstairs (Ходьба вниз), Sitting (Сидение), Standing (Стояние). Требуется построить алгоритм классификации, ставящий в соответствие временному ряду метку класса.

В качестве базового метода построения признакового пространства рассматривается метод, изложенный в работе [5]. Признаки задаются экспертным путем; для трехмерного временного ряда (X, Y, Z -компоненты ускорения) длины 200 (что соответствует 10 с) выбираются следующие 40 признаков:

- {3} — среднее ускорение по каждой оси;
- {3} — стандартное отклонение ускорения по каждой оси;
- {3} — среднее абсолютное отклонение ускорения по каждой оси;
- {1} — среднее результирующее ускорение;
- {30} — распределение значений временного ряда по каждой оси. Для каждой компоненты X, Y, Z вычисляется наибольшее и наименьшее значения на всем промежутке; область значений компоненты разбивается на 10 равных промежутков; для каждого промежутка вычисляется процент попавших в него значений компоненты временного ряда.

5 Вычислительный эксперимент

Приведем некоторые свойства признаков, получаемых оценкой параметров порождения временных рядов акселерометра. В этом разделе за критерий качества примем процент совпавших объектов внутри всех классов:

$$Q(\mathbf{b}) = \frac{1}{m} \sum_{i=1}^m I(g(\hat{\theta}_i, \mathbf{b}) = y_i). \quad (5)$$

5.1 Иллюстрация свойств моделей порождения

Согласно предыдущему разделу, моделью порождения временного ряда $a_i(t)$ будем считать модель авторегрессии с длиной предыстории n :

$$a_i(t) = \theta_0 + \sum_{j=1}^n \theta_j a_i(t-j) + \varepsilon_t. \quad (6)$$

Оптимальные параметры авторегрессионной модели $\hat{\theta}$ определим с помощью минимизации среднеквадратичной ошибки прогнозирования (4). На рис. 1 показаны примеры исходных временных рядов для каждого из шести классов, а также их авторегрессионный прогноз при оптимальных значениях параметров $\hat{\theta}_i$.

Для определения параметра n , оптимальной длины предыстории в модели авторегрессии, построим график зависимости критерия качества (5) от n на тестовой выборке. В качестве модели классификации рассмотрим модель многоклассовой логистической регрессии. График зависимости показан на рис. 2. Синей линией показан график среднего на контроле качества для модели одномерной авторегрессии (6), зеленой линией — для трехмерной, где авторегрессионные параметры вычисляются отдельно для каждой из компонент $\{a_x(t), a_y(t), a_z(t)\}_{t=1}^T$. Видно, что модель трехмерной авторегрессии существенно превосходит одномерную модель.

Для построения финальной модели классификации была выбрана трехмерная модель авторегрессии и длина предыстории для нее $n = 20$.

Иллюстрация собственных чисел для метода SSA показана на рис. 3: синие линии — примеры для класса Jogging; красные — Walking; зеленые — Standing. Для этих трех классов видна характерная делимость выборки по спектру.

5.2 Результаты классификации

В этом разделе приведем результаты классификации данных акселерометра для различных методов выделения признаков с использованием порядковой логистической регрессии и SVM в качестве классификаторов в новом признаковом пространстве.

В качестве процедуры разбиения выборки на обучение и контроль был реализован метод разбиения с повторениями. На каждой итерации выборка разбивалась случайно в пропорции 70% обучения / 30% контроль; разбиения происходили независимо 50 раз. Представленное в табл. 1 качество Q является значением средней достоверности (ассигасу) многоклассового классификатора на контрольных подвыборках:

$$Q = \frac{\sum_{k=1}^6 \text{tp}_k}{m},$$

где tp_k — количество правильно выделенных объектов внутри класса $k \in \{1 \dots 6\}$.

Отметим, что, во-первых, процедура разбиения с повторениями не позволяет оценить дисперсию исследуемой величины достоверности, однако позволяет с достаточной степенью точности провести оценку среднего значения достоверности в силу большого количества генерируемых подвыборок. Во-вторых, исследуемая величина достоверности имеет явный перекоп в сторону классов с большим количеством объектов, которыми в данной задаче являются классы Walking и Jogging. Для решения этой проблемы приводятся значения достоверности tp_k/m_k внутри каждого класса по отдельности.

Отметим также, что для итоговой классификации были выбраны два наиболее простых алгоритма классификации (многоклассовая логистическая регрессия и SVM с полиноми-

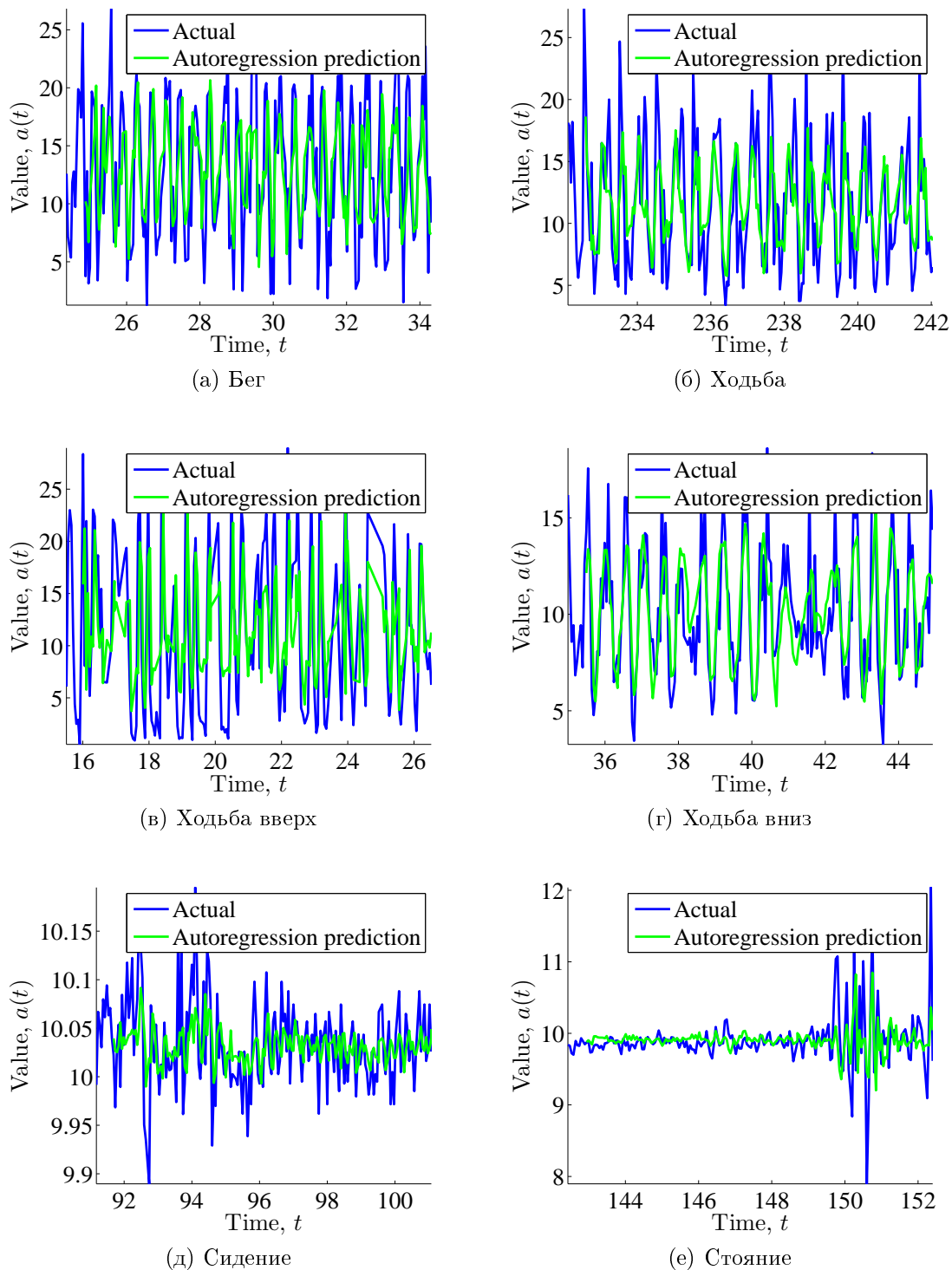


Рис. 1 Исходные данные и оптимальная авторегрессионная модель

альным ядром степени 3), разделяющих объекты в линейном пространстве гиперплоскостями простой формы. Такой выбор связан с желанием авторов продемонстрировать свойства нового признакового пространства, а не решающего правила классификации. Как будет видно из результатов эксперимента, даже такие алгоритмы классификации

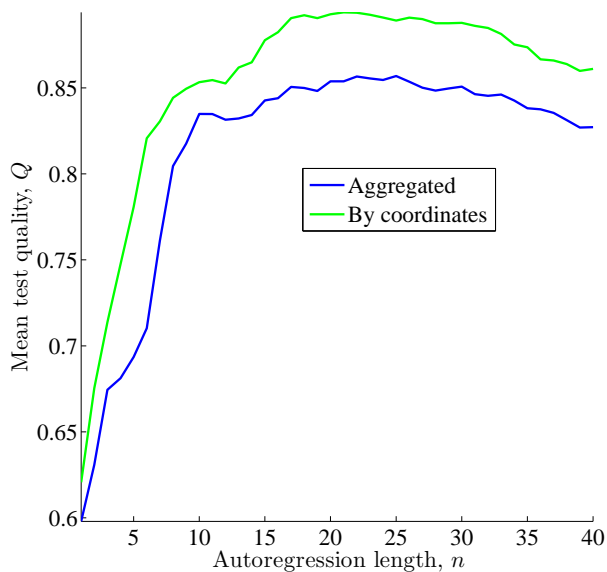


Рис. 2 Зависимость критерия качества от длины предыстории

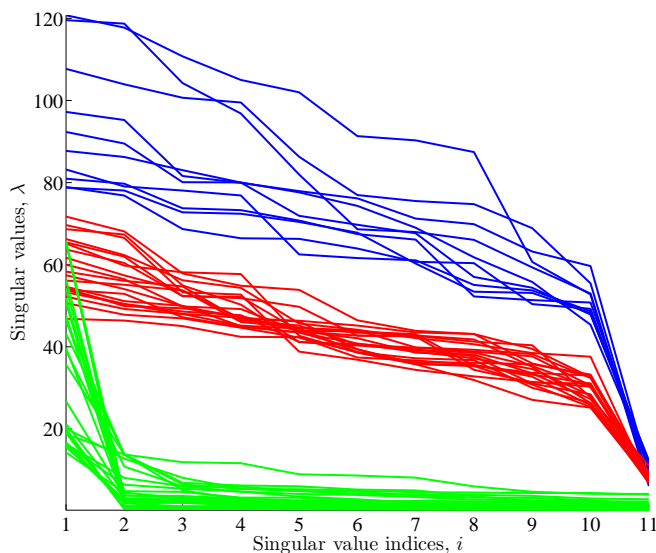


Рис. 3 Иллюстрация собственных чисел для метода SSA: синие линии — примеры для класса Бег; красные — Ходьба; зеленые — Стояние

позволяют достичь хороших результатов, что позволяет говорить о полноте множества выбранных признаков и переходить к решению задачи онлайн сегментации временного ряда по выбранным признакам.

Результаты классификации для модели авторегрессии с лагом $n = 20$ и логистической регрессии показаны на рис. 4, а. Синими столбиками отмечено общее количество объектов в классе, зелеными — среднее количество правильно классифицированных на контрольных подвыборках. Общее значение критерия качества составило $Q = 90\%$.

В табл. 1, а представлены результаты классификации. По горизонтали отложены актуальные значения классов, про вертикали — предсказанные. Число в таблице равно среднему количеству предсказаний за чужой класс. Последний столбец в таблице показывает долю правильно классифицированных объектов внутри каждого класса.

Таблица 1 Результаты классификации

(a) AR-модель, логистическая регрессия

	Предсказанный класс						
	Бег	Ходьба	Ходьба вверх	Ходьба вниз	Сидение	Стояние	Точность
Бег	469	6	4	5	0	0	0,97
Ходьба	3	604	6	10	1	1	0,97
Ходьба вверх	18	23	112	9	1	1	0,68
Ходьба вниз	10	15	19	86	2	2	0,64
Сидение	3	1	5	1	68	5	0,82
Стояние	2	0	3	2	5	58	0,83

(b) AR-GARCH-модель, SVM

	Предсказанный класс						
	Бег	Ходьба	Ходьба вверх	Ходьба вниз	Сидение	Стояние	Точность
Бег	480	4	3	1	0	0	0,98
Ходьба	2	614	4	6	0	1	0,98
Ходьба вверх	4	5	146	10	1	0	0,88
Ходьба вниз	2	4	11	115	0	1	0,86
Сидение	0	0	1	1	72	7	0,89
Стояние	0	1	1	1	8	57	0,84

(c) Анализ спектра, логистическая регрессия

	Предсказанный класс						
	Бег	Ходьба	Ходьба вверх	Ходьба вниз	Сидение	Стояние	Точность
Бег	475	10	3	5	1	0	0,96
Ходьба	2	595	12	6	1	2	0,96
Ходьба вверх	14	44	85	21	1	0	0,52
Ходьба вниз	1	34	28	68	0	0	0,52
Сидение	0	0	2	1	79	3	0,93
Стояние	0	5	0	2	2	60	0,87

(d) Ручное выделение признаков, SVM

	Предсказанный класс						
	Бег	Ходьба	Ходьба вверх	Ходьба вниз	Сидение	Стояние	Точность
Бег	483	4	4	2	0	0	0,98
Ходьба	1	607	10	9	0	0	0,97
Ходьба вверх	6	17	120	18	0	0	0,77
Ходьба вниз	1	10	12	106	0	0	0,82
Сидение	2	6	7	7	57	5	0,72
Стояние	0	9	8	4	2	45	0,71

(e) Совместное использование всех признаков, SVM

	Предсказанный класс						
	Бег	Ходьба	Ходьба вверх	Ходьба вниз	Сидение	Стояние	Точность
Бег	490	1	2	1	0	0	0,99
Ходьба	0	622	1	4	0	0	0,99
Ходьба вверх	1	2	154	5	0	0	0,95
Ходьба вниз	0	2	4	124	0	0	0,95
Сидение	0	1	2	0	79	1	0,95
Стояние	0	0	1	1	1	65	0,96

С помощью усложнения прогностической модели до AR-GARCH (с параметрами $AR = 20$, $ARCH = 2$, $GARCH = 1$), а также использования SVM с квадратичным ядром удалось увеличить качество классификации до 95% (см. рис. 4, б и табл. 1, б). Описание GARCH модели подробно изложено, например, в [6, гл. 14].

В качестве классифицирующей функции $g(\lambda_i, \mathbf{b})$ для спектрального метода выделения признаков рассмотрим многоклассовую логистическую регрессию. Результаты классификации спектрального метода показаны на рис. 4, в и в табл. 1, с. Несмотря на то что среднее качество на контрольной выборке составило 87%, следует отметить, что данный метод позволил лучше остальных выделить классы Sitting (93%) и Standing (87%).

На рис. 4, г и в табл. 1, д показаны результаты классификации для экспертного метода выделения признаков. Среднее качество на контрольной выборке составило 91%.

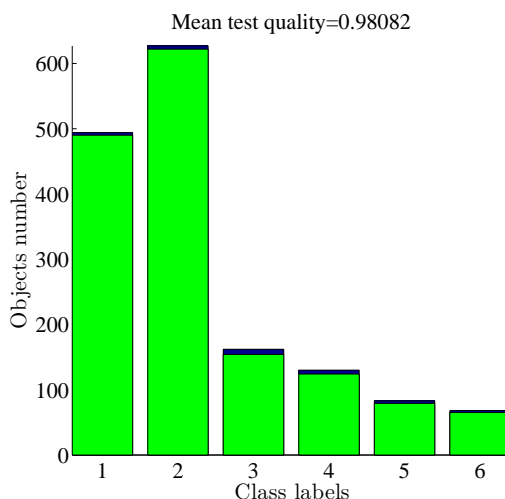
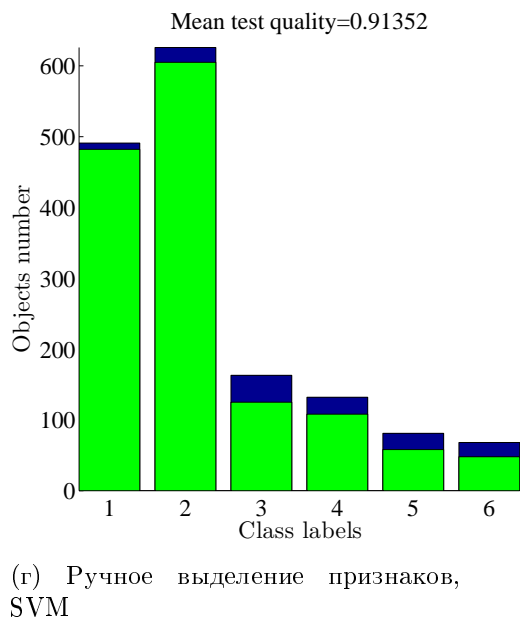
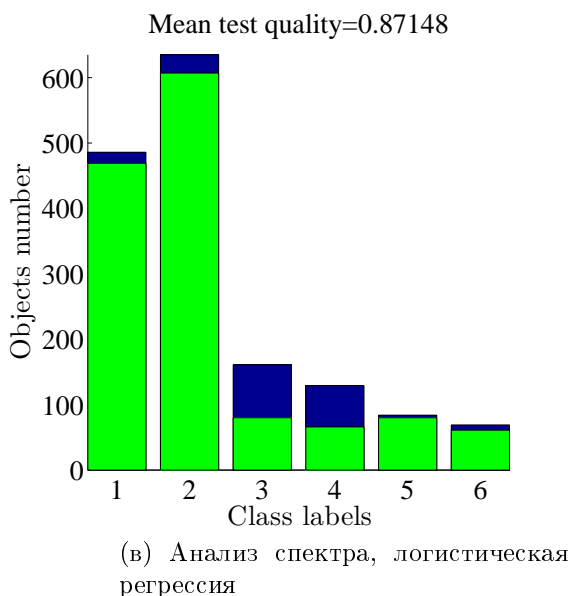
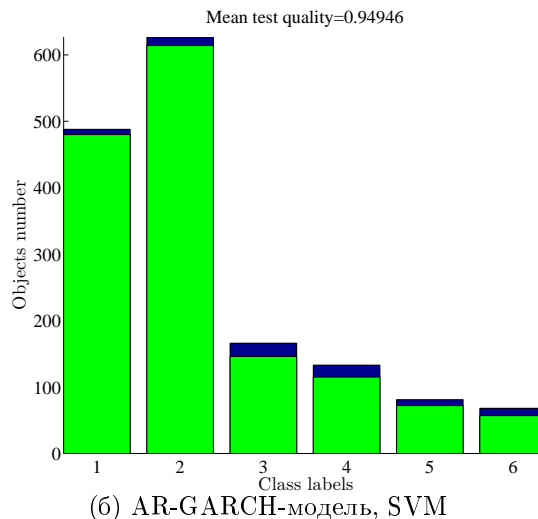
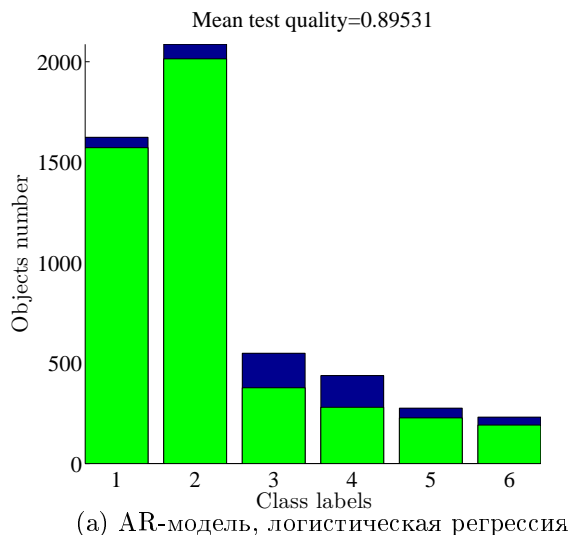


Рис. 4 Результаты классификации

Таблица 2 Итоговые результаты классификации для различных методов выделения признаков

Действие	Точность				
	AR-logistic	AR-Garch-SVM	Manual-SVM	Spectral-logistic	All feats-SVM
Бег	0,97	<i>0,98</i>	<i>0,98</i>	0,96	0,99
Ходьба	0,97	<i>0,98</i>	0,97	0,96	0,99
Ходьба вверх	0,68	<i>0,88</i>	0,75	0,52	0,95
Ходьба вниз	0,64	<i>0,86</i>	0,82	0,52	0,95
Сидение	0,82	0,89	0,68	<i>0,93</i>	0,95
Стояние	0,83	0,84	0,66	<i>0,87</i>	0,96
Суммарно	0,90	<i>0,95</i>	0,91	0,87	0,98

Результаты классификации для использования всех перечисленных признаков в совокупности и SVM с полиномиальным ядром степени 3 показаны на рис. 4, *d* и в табл. 1, *e*. Итоговое качество классификации составило 98%.

Итоговые результаты для всех методов представлены в табл. 2. Курсивом показаны наилучшие результаты классификации для отдельных методов выделения признаков. Видно, что наибольшую точность (95%) имеет метод AR-Garch-SVM, однако альтернативные методы выделения признаков лучше справляются с выделением отдельных классов. Результаты классификации в случае использования всех признаков в совокупности иллюстрирует последний столбец.

6 Заключение

Рассмотрена задача многоклассовой классификации временных рядов как объектов сложной структуры. Исследованы различные методы построения признаков и метод построения признаков на основе гипотезы порождения данных. Проведено исследование различных методов построения признаков и решающих правил в задаче классификации данных акселерометра. Результаты показывают, что построенное признаковое пространство достаточным образом описывает зависимую переменную и приводит к высоким результатам классификации.

Авторы выражают благодарность В. В. Стрижову за постановку задачи и внимательное отношение к работе.

Литература

- [1] *Geurts P.* Pattern extraction for time series classification // Principles of data mining and knowledge discovery. — Springer, 2001. P. 115–127.
- [2] *Wei L., Keogh E.* Semi-supervised time series classification // 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Proceedings. — New York, NY, USA: ACM, 2014. P. 748–753. <http://doi.acm.org/10.1145/1150402.1150498>.
- [3] *Wang W., Liu H., Yu L., Sun F.* Human activity recognition using smart phone embedded sensors: A linear dynamical systems method // Joint Conference (International) on Neural Networks (IJCNN), 2014. P. 1185–1190.
- [4] *Ignatov A., Strijov V.* Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer // Multimedia Tools Applications, 2015. P. 1–14. doi: 10.1007/s11042-015-2643-0. <http://dx.doi.org/10.1007/s11042-015-2643-0>.

- [5] *Kwapisz J. R., Weiss G. M., Moore S. A.* Activity recognition using cell phone accelerometers // *SIGKDD Explor. Newsl.*, 2011. Vol. 12. No.2. P. 74–82. doi: 10.1145/1964897.1964918. <http://doi.acm.org/10.1145/1964897.1964918>.
- [6] *Lukashin Y. P.* Adaptive methods for short-term forecasting. — *Finansy i Statistika*, 2003.
- [7] *Mörchen F.* Time series feature extraction for data mining using DWT and DFT. 2003.
- [8] *Zhang H., Ho T. B., Lin M. S.* A non-parametric wavelet feature extractor for time series classification // *Advances in knowledge discovery and data mining*. — Springer, 2004. P. 595–603.
- [9] *Hassani H.* Singular spectrum analysis: Methodology and comparison. — Cardiff University and Central Bank of the Islamic Republic of Iran, 2007. 19 p.

References

- [1] Geurts, P. 2001. Pattern extraction for time series classification. *Principles of data mining and knowledge discovery*. Springer. 115–127.
- [2] Wei, L., and E. Keogh. 2014. Semi-supervised time series classification. *12th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings*. New York, NY: ACM. P. 748–753. Available at: <http://doi.acm.org/10.1145/1150402.1150498>.
- [3] Wang, W., H. Liu, L. Yu, and F. Sun. 2014. Human activity recognition using smart phone embedded sensors: A linear dynamical systems method. *Joint Conference (International) on Neural Networks (IJCNN)*. 1185–1190.
- [4] Ignatov, A., and V. Strijov. 2015. Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. *Multimedia Tools Applications*. 1–14. doi: 10.1007/s11042-015-2643-0. Available at: <http://dx.doi.org/10.1007/s11042-015-2643-0>.
- [5] Kwapisz, J. R., G. M. Weiss, and S. A. Moore. 2011. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.* 12(2):74–82. doi: 10.1145/1964897.1964918. Available at: <http://doi.acm.org/10.1145/1964897.1964918>.
- [6] Lukashin, Y. P. 2003. *Adaptive methods for short-term forecasting*. *Finansy i Statistika*.
- [7] Mörchen, F. 2003. Time series feature extraction for data mining using DWT and DFT.
- [8] Zhang, H., T. B. Ho, and M. S. Lin. 2004. A non-parametric wavelet feature extractor for time series classification. *Advances in knowledge discovery and data mining*. Springer. 595–603.
- [9] Hassani, H. 2007. *Singular spectrum analysis: Methodology and comparison*. Cardiff University and Central Bank of the Islamic Republic of Iran.